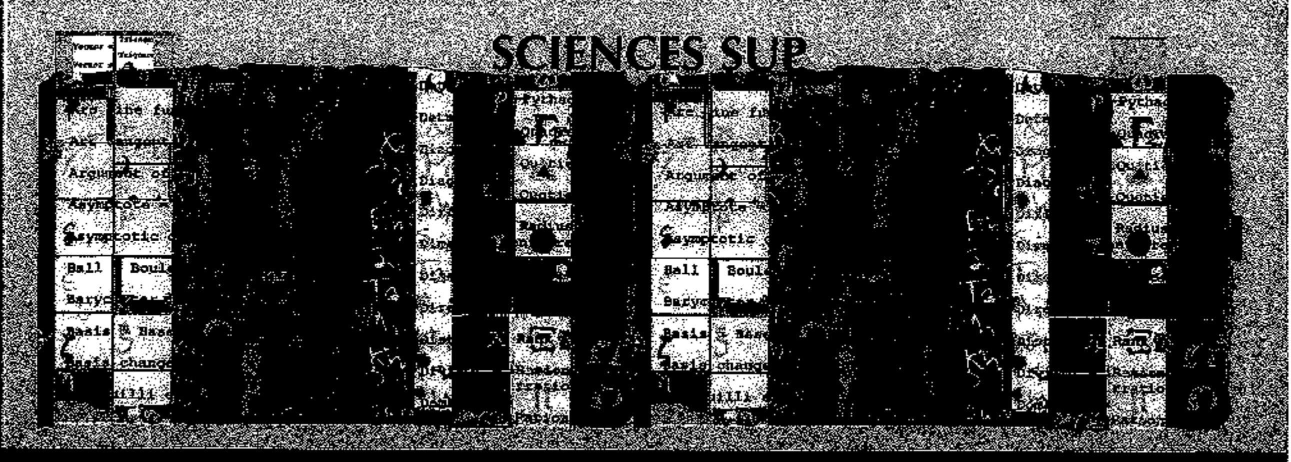


Cours et exercices corrigés

2e cycle/Master • Agrégation • Écoles d'ingénieurs

PROCESSUS STOCHASTIQUES

Processus de Poisson, chaînes de Markov et martingales

Dominique Foata
Aimé Fuchs



PROCESSUS STOCHASTIQUES

Processus de poisson, chaîne de Markov et martingales

PROCESSUS STOCHASTIQUES

Processus de poisson, chaînes de Markov et martingales

Cours et exercices corrigés

Domonique Foata
Professeur de mathématiques
à l'université Louis-Pasteur de Strasbourg

Aimé Fuchs
Professeur de mathématiques
à l'université Louis-Pasteur de Strasbourg



CHEZ LE MÊME ÉDITEUR

Calcul des probabilités, par Dominique Foata et Aimé Fuchs, 2e édition, 1998.

Illustration de couverture : *Lionel Auvergne*

ISBN 2 10 048850 3

Table des matières

Avant-propos

Ce cours de probabilités fait suite au livre *Calcul des Probabilités, cours, exercices et problèmes corrigés* (Dunod, Paris, seconde édition, 1998) écrits par les mêmes auteurs ; il présuppose, en principe, la connaissance des matières exposées dans ce précédent livre. Le présent cours reste, cependant, accessible à toute personne ayant acquis les connaissances de base du calcul des probabilités, telles qu'elles sont enseignées dans les cours de licence de mathématiques (bac+3) des Universités. Il sera fait naturellement référence au livre cité, dit volume 1, sous l'appellation « FF1 », mais les notations utilisées sont suffisamment classiques pour que tout lecteur préparé n'ait aucun mal à aborder la lecture de celui-ci. Un premier chapitre intitulé *Notations utilisées et rappels* a été inclus pour permettre une entrée en matières aisée.

Le présent cours s'adresse aux étudiants de maîtrise de mathématiques appliquées et d'informatique (bac+4), ainsi qu'aux élèves des grandes écoles d'ingénieurs qui s'orientent vers la recherche opérationnelle. L'esprit du précédent volume a été conservé : à la fin de chaque chapitre, à l'exclusion du premier qui contient les rappels et de l'avant-dernier qui contient des exemples d'applications, on trouve un ensemble d'exercices, qui reçoivent des solutions souvent détaillées.

Ce cours a pris corps à l'Université Louis Pasteur, dans le programme du magistère d'actuariat (bac +3 à +5), lorsque chacun des auteurs, à tour de rôle, a assuré les cours de probabilités. Une fois acquises les connaissances exposées dans FF1, au cours de la première année et du premier semestre de la deuxième année, les étudiants engagés dans ce programme s'initient aux éléments de la théorie des processus stochastiques, durant le second semestre

de la deuxième année et le premier semestre de la troisième. Nous avons donc conçu un cours reposant sur trois processus très classiques : les processus de Poisson, les chaînes de Markov à valeurs discrètes, les martingales à temps discret. Nous terminons l'ouvrage par une étude sur le mouvement brownien.

Naturellement, processus de Poisson et chaînes de Markov ont fait l'objet de nombreux exposés. On les trouve traités dans les cours de Neveu [8] et Carton [2]. Il existe aussi un excellent ouvrage sur le seul processus de Poisson par Kingman [5]. La théorie des chaînes de Markov est traitée dans le classique ouvrage de Feller [4] et dans bien d'autres livres plus anciens.

La théorie des martingales à temps discret est traitée de façon exhaustive par Neveu [7]. Voir aussi le livre pétillant de Williams [9] ou le très classique manuel de Bauer [1]. Comme cette théorie est plus récente (elle a quand même soixante ans!), les exercices sont moins classiques. Nous avons, pour cette partie, bénéficié des précieux conseils de notre collègue Giorgio Letta. Mentionnons, enfin, le traité de Loève [6], avec son chapitre sur le mouvement brownien et aussi le superbe livre d'exercices rédigé par les probabilistes ukrainiens [3].

Le premier chapitre contient un rappel sur les *règles de calcul* des espérances conditionnelles, auquel nous nous sommes constamment reportés pour le traitement de la théorie des martingales dans les chapitres 7–11. Le livre se termine par un chapitre traitant d'exemples d'applications et par une brève étude sur le mouvement brownien. Comme ce processus est désormais un ingrédient fréquent dans certains modèles de mathématiques financières, nos étudiants ont accueilli avec sympathie un tel chapitre dans notre ouvrage.

Beaucoup d'exercices sont classiques. Certains sont nouveaux, ou rédigés de façon nouvelle. Une grande partie d'entre eux ont été testés auprès des étudiants du Magistère d'actuariat. Nous les remercions de nous avoir fait part de leurs critiques. Nous sommes, encore une fois, reconnaissants à Anatole Joffe pour toutes ses remarques bienveillantes et érudites, ainsi qu'à Michel Émery et Jacques Franchi pour leurs commentaires éclairés, notamment sur le mouvement brownien.

Comme le souhaitait l'éditeur, le texte a été mis, par les auteurs, en conformité avec la charte de présentation des Éditions Dunod. Les théorèmes, propositions, propriétés ont reçu une numérotation absolue. Par exemple, le « Théorème 8.2.1 » est le premier théorème apparaissant dans le deuxième paragraphe du chapitre 8. Les définitions également sont pourvues d'une telle numérotation. En revanche, les remarques sont le plus souvent mentionnées sans aucune numérotation.

Les auteurs voudraient remercier Vanessa Beunèche, des Éditions Dunod, pour sa lecture attentive du texte et pour ses remarques avisées, qui ont amélioré la présentation typographique finale. Ils sont, enfin, reconnaissants à André Violante, qui a leur a fait part de toutes ses innovations en matière de typographie informatique.

Dominique FOATA
Aimé FUCHS
Département de mathématique
Université Louis Pasteur
7, rue René-Descartes,
F-67084 Strasbourg

courriel : `foata@math.u-strasbg.fr`

BIBLIOGRAPHIE

[1] Bauer (Heinz). — *Probability Theory.* — Walter de Gruyter & Co., Berlin, 1996.
[2] Carton (Daniel). — *Processus aléatoires utilisés en recherche opérationnelle.* — École Centrale de Paris, 1993.
[3] Dogorovtsev (A. Ya.), Silvestrov (D. S.), Skorokhod (A. V.), Yadrenko (M. I.). — *Probability Theory : Collection of Problems.* — Amer. Math. Soc., Translations of Math. Monographs, vol. 163, Providence, R.I., 1991.
[4] Feller (William). — *An Introduction to Probability and its Applications,* vol. 1, 3rd Edition. — John Wiley, New York, 1968.
[5] Kingman (J. F. C.). — *Poisson Processes.* — Clarendon Press, Oxford, 1993.
[6] Loève (Michel). — *Probability Theory II,* 4th Edition. — Springer-Verlag, New York, 1978.
[7] Neveu (Jacques). — *Martingales à temps discret.* — Masson, Paris, 1972.
[8] Neveu (Jacques). — *Introduction aux probabilités.* — École Polytechnique, Paris, 1994.
[9] Williams (David). — *Probability with martingales.* — Cambridge Univ. Press, 1991.

Chapitre Premier

Notations utilisées et Rappels

Le but de ce premier chapitre est d'exposer rapidement les notions de base des probabilités, telles qu'elles ont été développées, par exemple, dans le livre publié précédemment par les mêmes auteurs[1] et de préciser les notations qui seront constamment utilisées dans le présent ouvrage. À la fin de chaque paragraphe, est donnée une référence au livre cité, sous l'appellation «FF1». Le chapitre se termine par un exposé de quelques propriétés des tribus engendrées par des variables aléatoires (§ 12) et par un traitement des *espérances conditionnelles* (§ 13), dont l'usage est constant dans la théorie des *martingales*. On imagine que le lecteur fera une première lecture rapide de ce chapitre, quitte à y revenir pour s'assurer d'une notation ou d'une propriété sur les espérances conditionnelles.

1.1 LE TRIPLET FONDAMENTAL

Il est noté $(\Omega, \mathfrak{A}, \mathrm{P})$ et consiste en un ensemble non-vide Ω, un ensemble $\mathfrak{A}$ de parties de Ω, muni d'une structure de *tribu* et une *mesure de probabilité* P sur la tribu $\mathfrak{A}$. Les éléments de Ω sont les *épreuves* attachées au problème sous-jacent et les éléments de la tribu $\mathfrak{A}$ sont les *évènements* auxquels on peut attacher une *probabilité*. Le triplet est appelé *espace probabilisé*.

Les opérations ensemblistes, comme la réunion et l'intersection, conservent leurs notations classiques. Le *complémentaire* d'un évènement A est noté A^c.

[1] Foata (Dominique) & Fuchs (Aimé). — *Calcul des probabilités; cours, exercices et problèmes corrigés*, 2ième édition. — Paris, Dunod, 1998.

Si les évènements A et B sont *disjoints* (on dit encore *incompatibles*), on écrit aussi $A+B$ pour la réunion $A \cup B$. La *différence* $A \cap B^c$ entre l'évènement A et l'évènement B est notée $A \setminus B$.

Pour une suite (A_n) d'évènements, on construit les évènements : $\bigcup_n A_n$ « au moins l'un des A_n se réalise » ; $\bigcap_n A_n$ « tous les A_n se réalisent » ; $\liminf_n A_n$ « tous les A_n se réalisent après un certain indice » ; $\limsup_n A_n$ « une infinité d'évènements A_n se réalisent ». Si les évènements A_n sont disjoints deux à deux, on écrit aussi $\sum_n A_n$ pour la réunion $\bigcup_n A_n$. À chaque évènement A, on associe son *indicatrice*, notée I_A, qui est la fonction, définie sur Ω, valant 1 sur A et 0 ailleurs.

La tribu *engendrée par* une famille $\mathcal{C}$ d'évènements est notée $\mathfrak{T}(\mathcal{C})$. C'est la plus petite tribu, pour la relation d'inclusion, qui contient $\mathcal{C}$. La *tribu borélienne* sur $\mathbb{R}$ est notée $\mathcal{B}^1$. Elle est, par exemple, engendrée par la classe des intervalles fermés bornés. La *tribu borélienne* de $\mathbb{R}^n$ est notée $\mathcal{B}^n$. On peut la définir comme la tribu engendrée par la famille des *pavés* bornés, à savoir les produits d'intervalles fermés bornés : $\prod_{1\le i\le n} [a_i, b_i]$.

Voir [FF1], chap. 1, 2 et 3.

1.2 LE THÉORÈME BINOMIAL

Ce théorème intervient si souvent dans les calculs explicites, qu'il paraît important de rappeler son énoncé. Si n est un entier positif et a un élément de $\mathbb{C}$, la *factorielle montante* est définie par

$$(a)_n := \begin{cases} 1, & \text{si } n = 0\,; \\ a(a+1)\cdots(a+n-1), & \text{si } n \geq 1. \end{cases}$$

La fonction *hypergéométrique*, dépendant d'une suite de p paramètres $a_1, \ldots,$ a_p et d'une suite de q paramètres $b_1, \ldots, b_q$ est la série de puissances en x :

$${}_pF_q\Big(\begin{matrix} a_1, \ldots, a_p \\ b_1, \ldots, b_q \end{matrix}; x\Big) := \sum_{n\geq 0} \frac{(a_1)_n \cdots (a_p)_n}{(b_1)_n \cdots (b_q)_n} \frac{x^n}{n!}.$$

Lorsque $p = q = 0$, la fonction hypergéométrique est simplement la fonction exponentielle

$${}_0F_0\Big(\begin{matrix} - \\ - \end{matrix}; x\Big) = \sum_{n\geq 0} \frac{x^n}{n!} = \exp x.$$

Lorsque $p = 1$ et $q = 0$, la fonction hypergéométrique peut être sommée pour tout x tel que $|x| < 1$. On obtient l'identité du *théorème binomial* :

$${}_1F_0\Big(\begin{matrix} a \\ - \end{matrix}; x\Big) = \sum_{n\geq 0} (a)_n \frac{x^n}{n!} = (1-x)^{-a}.$$

Le *coefficient binomial* $\dbinom{a}{n}$ (où a est un complexe et n un entier positif) est

défini par : $\binom{a}{n} := \dfrac{a(a-1)\cdots(a-n+1)}{n!} = (-1)^n \dfrac{(-a)_n}{n!}$.

L'*identité de Chu-Vandermonde* s'exprime par :

$$ {}_2F_1\left(\begin{matrix}-n, a\\ c\end{matrix}; 1\right) = \frac{(c-a)_n}{(c)_n}. $$

Voir [FF1], chap. 3.

1.3 CONDITIONNEMENT ET INDÉPENDANCE

Si A est un évènement de probabilité $\mathrm{P}(A) > 0$, la mesure de probabilité *conditionnelle, A étant donné*, associe à tout évènement B la probabilité $\mathrm{P}(B \mid A) := \mathrm{P}(B \cap A)/P(A)$. La formule du *conditionnement en chaîne* s'exprime en disant que si $n \geq 2$ et si $A_1, A_2, \ldots, A_n$ est une suite de n évènements tels que $\mathrm{P}(A_1A_2\ldots A_{n-1}) > 0$, on a l'identité :

$$ (3.1) \quad \mathrm{P}(A_1A_2\ldots A_n) = \mathrm{P}(A_n \mid A_1\ldots A_{n-1})\mathrm{P}(A_{n-1} \mid A_1\ldots A_{n-2})\cdots\mathrm{P}(A_2 \mid A_1)\mathrm{P}(A_1). $$

On dit qu'une *suite* (A_n) d'évènements est un *système complet* si

(i) $i \neq j \Rightarrow A_i \cap A_j = \emptyset$ (les évènements sont deux à deux incompatibles) ;

(i) $\mathrm{P}(\sum_n A_n) = \sum_n \mathrm{P}(A_n) = 1$ (presque sûrement l'un des évènements A_n se réalise).

Deux évènements A et B sont *indépendants* si l'on a $\mathrm{P}(A\cap B) = \mathrm{P}(A)\mathrm{P}(B)$. Soit (X_n) une suite de variables aléatoires. On dit que c'est une suite de variables (mutuellement) *indépendantes*, si pour toute suite finie $(i_1 < \cdots < i_r)$ d'entiers telle que $r \geq 2$ et toute suite $(B_1, \ldots, B_r)$ d'ensembles boréliens, on a l'identité : $\mathrm{P}\{X_{i_1} \in B_1, \ldots, X_{i_r} \in B_r\} = \mathrm{P}\{X_{i_1} \in B_1\}\cdots\mathrm{P}\{X_{i_r} \in B_r\}$.

Voir [FF1], chap. 6.

1.4 PROBABILITÉS DISCRÈTES ET VARIABLES ALÉATOIRES

La mesure de probabilité *singulière* en ω_0 est notée ε_{ω_0}. Pour chaque évènement A, on a $\varepsilon_{\omega_0}(A) = 1$ ou 0, suivant que A contient ω_0 ou non. Si (α_n) est une suite de nombres positifs tels que la somme de la série $\sum_n \alpha_n$ vaut 1 et si (ω_n) est une suite d'éléments de Ω, la somme $\sum_n \alpha_n \varepsilon_{\omega_n}$ est une mesure de probabilité *discrète*, portée par la suite (ω_n).

On utilise cette notation pour exprimer les lois de probabilité discrètes usuelles, comme la *loi binomiale* $B(n,p) := \sum_k \binom{n}{k} p^k q^{n-k} \varepsilon_k$ (avec $q = 1-p$) ou la *loi de Poisson* $\pi_\lambda := \sum_{k\geq 0} e^{-\lambda}(\lambda^k/k!)\,\varepsilon_k$, ou encore la *loi géométrique*, qui apparaît sous les deux formes $\sum_{k\geq 1} pq^{k-1}\,\varepsilon_k$ et $\sum_{k\geq 0} pq^k\,\varepsilon_k$.

Les *variables aléatoires* sont notées X, Y, … avec ou sans indices. Si elles sont définies sur l'espace probabilisé $(\Omega, \mathfrak{A}, \mathrm{P})$ et si B est un ensemble borélien, alors $\{X \in B\}$ est l'évènement $\{\omega \in \Omega : X(\omega) \in B\}$. La *loi de probabilité* de X

est notée P_X ; elle est définie, pour tout $B \in \mathcal{B}^1$, par $\mathrm{P}_X(B) := \mathrm{P}(X^{-1}(B)) = \mathrm{P}\{X \in B\}$. La *fonction de répartition* de X est la fonction de variable réelle, à valeurs dans $[0,1]$, définie par : $\mathrm{F}_X(x) := \mathrm{P}\{X \leq x\}$.

Si la loi de probabilité de X est discrète, soit $\mathrm{P}_X = \sum_n \alpha_n \varepsilon_{x_n}$ et si la série $\sum_n \alpha_n |x_n|$ est convergente, l'*espérance mathématique* de X est définie par $\mathbb{E}[X] := \sum_n \alpha_n x_n$.

Si X est une variable aléatoire prenant ses valeurs dans $\mathbb{N} = \{0, 1, \dots\}$, sa *fonction génératrice* est la série de puissances $G_X(s) := \sum_n \mathrm{P}\{X = n\} s^n$. Si la série $G_X(s)$ est dérivable à gauche en $s = 1$, on a : $\mathbb{E}[X] = G'_X(1)$. Si $(N, X_1, X_2, \dots)$ est une suite de variables aléatoires indépendantes, toutes à valeurs dans $\mathbb{N}$, si les X_i ont même loi que la variable aléatoire X, alors la fonction génératrice de la *somme aléatoire* $S_N := X_1 + \cdots + X_N$ est donnée par $G_{S_N}(s) = G_N(G_X(s))$.

Voir [FF1], chap. 6–9.

1.5 INTÉGRATION

On définit l'*espérance mathématique* $\mathbb{E}[X]$ d'une variable aléatoire réelle X, comme l'intégrale $\int X\, d\mathrm{P}$, si elle existe, de X par rapport à la mesure P, de sorte que les propriétés de l'espérance mathématique résultent des propriétés de l'intégrale.

Soit f une fonction mesurable réelle d'une variable réelle. Le *théorème de transfert* permet de « transférer » le calcul de l'intégrale de $f \circ X$ par rapport à P, en un calcul d'intégrale sur $\mathbb{R}$, par rapport à la mesure de probabilité P_X. On a : $\int_\Omega f \circ X\, d\mathrm{P} = \int_\mathbb{R} f\, d\mathrm{P}_X$; en particulier, $\mathbb{E}[X] = \int_\mathbb{R} x\, d\mathrm{P}_X(x)$, avec les conventions habituelles sur les convergences d'intégrales.

Soit f une fonction réelle, positive, d'une variable réelle, d'intégrale de Lebesgue égale à 1 sur $\mathbb{R}$. On dit qu'une variable aléatoire réelle est *absolument continue, de densité* f, si pour tout ensemble borélien B, on a $\mathrm{P}_X(B) = \int_B f(x)\, dx$. Soit g une fonction (mesurable) réelle d'une variable réelle. Alors $\int_\mathbb{R} g(x)\, d\mathrm{P}_X(x) = \int_\mathbb{R} g(x) f(x)\, dx$, avec les conventions habituelles sur les convergences d'intégrales.

Soient X et Y deux variables aléatoires réelles, indépendantes, de fonctions de répartition respectives F et G. Alors la somme $X + Y$ a pour fonction de répartition le *produit de convolution* de F et G, donné pour tout réel z par :

$$(F * G)(z) := \int_\mathbb{R} \mathrm{F}(z-y)\, d\mathrm{G}(y).$$

Si X et Y sont absolument continues de densités respectives f et g, alors $X + Y$ a une densité donnée par le produit de convolution

$$(f * g)(z) = \int_\mathbb{R} f(z-y) g(y)\, dy.$$

Voir [FF1], chap. 10–11.

1.6 VECTEURS ALÉATOIRES

Soit (X, Y) un couple de variables aléatoires réelles définies sur un espace probabilisé $(\Omega, \mathfrak{A}, \mathrm{P})$. La loi de probabilité de (X, Y) est définie, pour tout borélien $B \in \mathcal{B}^2$, par $\mathrm{P}_{(X,Y)}(B) := \mathrm{P}\{(X, Y) \in B\}$. La loi du couple détermine complètement les lois de chacune des variables X et Y.

On dit que le couple (X, Y) est *absolument continu, de densité* $f_{(X,Y)}$, si $f_{(X,Y)}$ est une fonction réelle, de deux variables réelles, d'intégrale égale à 1 sur $\mathbb{R}^2$ et si, pour tout borélien $B \in \mathcal{B}^2$, on a : $\mathrm{P}_{(X,Y)}(B) = \int_B f_{(X,Y)}(x, y)\, dxdy$. Soit $\mathrm{F}_{(X,Y)}(x, y)$ la fonction de répartition de (X, Y), égale à $\mathrm{P}\{X \leq x, Y \leq y\}$. Alors, si $f_{(X,Y)}$ est continue en (x_0, y_0), on a : $f_{(X,Y)}(x_0, y_0) = (\partial^2/(\partial x\, \partial y)\mathrm{F}_{(X,Y)}(x, y)$ au point $(x, y) = (x_0, y_0)$.

Les densités de X et Y sont données par :

$$f_X(x) = \int_{\mathbb{R}} f_{(X,Y)}(x, y)\, dy, \qquad f_Y(y) = \int_{\mathbb{R}} f_{(X,Y)}(x, y)\, dx.$$

Les variables aléatoires X et Y sont indépendantes si et seulement si l'on a $f_{(X,Y)}(x, y) = f_X(x) f_Y(y)$ à un ensemble de mesure (de Lebesgue) près sur $\mathbb{R}^2$. Le rappel sur les *densités conditionnelles* est fait dans le paragraphe 13.

Voir [FF1], chap. 12.

1.7 FONCTIONS GÉNÉRATRICE ET CARACTÉRISTIQUE

Soit X une variable aléatoire réelle. Si la fonction réelle $g_X : u \mapsto \mathbb{E}[e^{uX}]$ est définie dans un voisinage ouvert de l'origine, on appelle *fonction génératrice des moments* de X la fonction g_X. Dans un voisinage ouvert de l'origine, on a alors :

$$g_X(t) = 1 + \mathbb{E}[X]\frac{t}{1!} + \mathbb{E}[X^2]\frac{t^2}{2!} + \cdots + \mathbb{E}[X^n]\frac{t^n}{n!} + \cdots$$

La *fonction caractéristique* $\varphi_X(t)$ de X est définie par : $\varphi_X(t) := \mathbb{E}[e^{itX}]$. Si X et Y sont deux variables aléatoires *indépendantes*, on a alors : $\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t)$. Si $\mathbb{E}[|X|]$ est finie, alors φ_X est continûment dérivable et l'on a : $\varphi'_X(0) = i\,\mathbb{E}[X]$. Enfin, si deux variables aléatoires ont même fonction caractéristique, elles ont même loi de probabilité.

Voir [FF1], chap. 13.

1.8 LOIS NORMALES ET EXPONENTIELLES, LOI GAMMA

Une variable aléatoire, de densité

$$\frac{1}{\sqrt{2\pi}} e^{-x^2/2} \quad (x \in \mathbb{R})$$

est dite *normale* (centrée, réduite). La loi de cette variable est appelée *loi normale réduite*; elle est notée $N(0, 1)$.

La *loi normale* $N(\mu, \sigma)$, où μ est un nombre réel et σ un nombre strictement positif, est la loi de densité

$$\frac{1}{\sqrt{2\pi}\,\sigma} \exp\Bigl(-\frac{1}{2}\,\frac{(x-\mu)^2}{\sigma^2}\Bigr) \quad (x \in \mathbb{R}).$$

Une variable aléatoire X, dont la loi de probabilité est la loi normale $N(\mu, \sigma)$ est d'*espérance mathématique* $\mathbb{E}[X] = \mu$ et de *variance* $\operatorname{Var} X = \sigma^2$.

Soient $n \geq 1$ et ${}^tX = (X_1, \dots, X_n)$ un vecteur aléatoire, dont les composantes X_i ont des moments du second ordre. (On note X le vecteur-*colonne* transposé de ${}^tX$.) Le nombre $\operatorname{Cov}(X_i, X_j) := \mathbb{E}[(X_i - \mathbb{E}[X_i])(X_j - \mathbb{E}[X_j])]$ est la *covariance* de X_i et X_j, si $i \neq j$ et la *variance*, encore notée $\operatorname{Var} X_i$, si $i = j$. La matrice $\Gamma = (\operatorname{Cov}(X_i, X_j))$ $(1 \leq i, j \leq n)$ est appelée *matrice des variances-covariances.* On dit qu'un vecteur aléatoire ${}^tX = (X_1, \dots, X_n)$ suit la *loi normale centrée* $\mathcal{N}(0, \Gamma)$, s'il admet pour fonction génératrice $g_X(u) = \exp\Bigl(\frac{1}{2}\,{}^tu\Gamma u\Bigr)$, où ${}^tu := (u_1, \dots, u_n)$. Soit ${}^t\mu = (\mu_1, \dots, \mu_n)$ un vecteur de $\mathbb{R}^n$. On dit que X suit la loi normale $\mathcal{N}(\mu, \Gamma)$, si $X - \mu$ suit la loi $\mathcal{N}(0, \Gamma)$. Si Γ est une matrice régulière, alors le vecteur X admet la densité

$$f_X(x) = \frac{1}{(2\pi)^{n/2} \det \Gamma} \exp\Bigl(-\frac{1}{2}\,{}^t(x-\mu)\,\Gamma^{-1}\,(x-\mu)\Bigr),$$

où ${}^tx := (x_1, \dots, x_n)$. Lorque $n = 1$, on utilise les deux notations $N(0, \sigma)$ et $\mathcal{N}(0, \sigma^2)$, pour la loi normale centrée, d'écart-type σ (et donc de variance σ^2).

Soit λ un nombre strictement positif; une variable aléatoire X à valeurs dans $]0, +\infty[$ est dite *exponentielle de paramètre* λ si elle est absolument continue et admet pour densité

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{si } x \geq 0\,; \\ 0, & \text{sinon.} \end{cases}$$

La loi de densité f est appelée *loi exponentielle* de *paramètre* λ $(\lambda > 0)$ et est notée $\mathcal{E}(\lambda)$. Sa *fonction de survie* vaut $r(x) = e^{-\lambda x}$ $(x > 0)$.

Une variable aléatoire est dite suivre la *loi* $\Gamma(p, \lambda)$ (*gamma de paramètres* $p > 0$, $\lambda > 0$) si elle est absolument continue et admet pour densité :

$$f(x) = \begin{cases} \dfrac{\lambda}{\Gamma(p)} e^{-\lambda x} (\lambda x)^{p-1}, & \text{si } x \geq 0\,; \\ 0, & \text{sinon.} \end{cases}$$

La somme de n variables aléatoires indépendantes, suivant chacune une loi exponentielle de paramètre λ, suit une loi $\Gamma(n, \lambda)$.

Voir [FF1], chap. 14, chap. 12.

1.9 CONVERGENCES

Soit (X_n) une suite de variables aléatoires réelles définies sur un espace probabilisé $(\Omega, \mathfrak{A}, \mathrm{P})$. On désigne par $(\mathbf{F}_n)$ la suite des fonctions de répartition

correspondantes. Soit X une variable aléatoire réelle, définie sur le même espace, de fonction de répartition F. On dit que la suite (X_n) *converge en loi* vers X et l'on écrit $X_n \xrightarrow{\mathcal{L}} X$, si $\mathrm{F}_n(x) \to \mathrm{F}(x)$ en tout point de continuité de F.

On dit que la suite *converge en probabilité vers* X et l'on écrit $X_n \xrightarrow{p} X$, si, pour tout $\varepsilon > 0$, on a : $\lim_n \mathrm{P}\{|X_n - X| > \varepsilon\} = 0$. On dit que la suite *converge presque sûrement vers* X et l'on écrit $X_n \xrightarrow{p.s.} X$, si $\mathrm{P}\{\lim_n X_n = X\} = 1$. La convergence presque sûre entraîne la convergence en probabilité, qui elle-même entraîne la convergence en loi.

Voir [FF1], chap. 16.

1.10 LOIS DES GRANDS NOMBRES

Soit (X_n) une suite de variables aléatoires, *centrées*, définies sur le même espace probabilisé. On dit que la suite (X_n) satisfait la *loi faible des grands nombres* si la suite de terme général $(1/n)\sum_{k=1}^{n} X_k$ converge vers 0 *en probabilité*. La suite (X_n) satisfait la *loi forte des grands nombres* si la suite de terme général $(1/n)\sum_{k=1}^{n} X_k$ converge vers 0 *presque sûrement*.

Soit (X_n) une suite de variables aléatoires, centrées, indépendantes deux à deux (resp. mutuellement indépendantes) et identiquement distribuées. Alors la suite (X_n) satisfait la loi faible des grands nombres (resp. la loi forte des grands nombres).

Soit (A_n) une suite d'évènements. Le *lemme de Borel-Cantelli* affirme que si la *série* de terme général $\mathrm{P}(A_n)$ est convergente, alors $\mathrm{P}(\limsup_n A_n) = 0$. Si les évènements A_n sont indépendants deux à deux et si la série de terme général $\mathrm{P}(A_n)$ est divergente, alors $\mathrm{P}(\limsup_n A_n) = 1$.

Voir [FF1], chap. 17.

1.11 LE THÉORÈME « CENTRAL LIMIT »

Soit (X_n) une suite de variables aléatoires indépendantes, équidistribuées ayant un moment du second ordre. On pose $\mu := \mathbb{E}[X_1]$, $\sigma^2 := \operatorname{Var} X_1$ et $Y_n := ((X_1 - \mu) + \cdots + (X_n - \mu))/(\sigma\sqrt{n})$. Le théorème « central limit » affirme alors que la suite (Y_n) converge en loi vers une variable aléatoire normale centrée réduite.

Voir [FF1], chap. 18.

1.12 TRIBU ENGENDRÉE PAR DES VARIABLES ALÉATOIRES

Soit Y une variable aléatoire à n dimensions, définie sur un espace probabilisé $(\Omega, \mathfrak{A}, \mathrm{P})$. La tribu $\mathfrak{T}(Y)$, engendrée par Y, est définie comme étant l'image réciproque $Y^{-1}(\mathcal{B}^n)$ de la tribu borélienne $\mathcal{B}^n$. C'est aussi la tribu engendrée

par les évènements de la forme $\{Y \in B\}$, où $B \in \mathcal{B}^n$. C'est encore la plus petite tribu, contenue dans $\mathfrak{A}$, qui rend Y mesurable (*cf.* [FF1], chap. 5, § 7).

Soit g une fonction *réelle* définie sur Ω. La proposition suivante fournit une condition nécessaire et suffisante pour que g soit mesurable par rapport au couple $(\Omega, \mathfrak{T}(Y))$ c'est-à-dire pour qu'elle satisfasse : $g^{-1}(\mathcal{B}^1) \subset \mathfrak{T}(Y)$.

Proposition 1.12.1. *Conservons les mêmes notations que ci-dessus. Alors la fonction g est mesurable par rapport au couple $(\Omega, \mathfrak{T}(Y))$ si et seulement s'il existe une fonction mesurable f de $(\mathbb{R}^n, \mathcal{B}^n)$ dans $(\mathbb{R}, \mathcal{B}^1)$ telle que $g = f \circ Y$. En particulier, l'indicatrice I_A d'un évènement A qui appartient à $\mathfrak{T}(Y)$ est de la forme $f \circ Y$.*

Démonstration. D'abord, si $g = f \circ Y$, pour tout $B \in \mathcal{B}^1$, on a $g^{-1}(B) = Y^{-1}(f^{-1}(B))$. Comme $f^{-1}(B) \in \mathcal{B}^n$, on a bien $g^{-1}(B) \in Y^{-1}(\mathcal{B}^n) = \mathfrak{T}(Y)$.

Pour la réciproque, on utilise le schéma traditionnel, (*cf.* [FF1], chap. 10, § 5) pour établir de telles propriétés. On vérifie la propriété pour les variables aléatoires simples et positives, puis pour les variables aléatoires positives, enfin pour les variables aléatoires quelconques.

Si g est une variable aléatoire réelle par rapport au couple $(\Omega, \mathfrak{T}(Y))$, à tout $B \in \mathcal{B}^1$ correspond un $C \in \mathcal{B}^n$ tel que $g^{-1}(B) = Y^{-1}(C)$. Soit $g = \sum_{1 \le k \le m} x_k I_{A_k}$ une variable aléatoire simple positive par rapport à $(\Omega, \mathfrak{T}(Y))$, telle que $\{A_1, \dots, A_m\}$ soit une partition de Ω en éléments de $\mathfrak{A}$. Il lui correspond une partition $\{C_1, \dots, C_m\}$ de $Y(\Omega)$ telle que, pour tout $k = 1, 2, \dots, m$, on ait $C_k \in \mathcal{B}^n$ et $g^{-1}(\{x_k\}) = A_k = Y^{-1}(C_k)$. On pose alors $f := \sum_{1 \le k \le m} x_k I_{C_k}$. On a bien $g = f \circ Y$; de plus, $f(y) = 0$ pour tout $y \in \mathbb{R}^n \setminus Y(\Omega)$.

Si g est une variable aléatoire positive, alors $g = \sup_n g_n$, où (g_n) est une suite croissante de variables aléatoires simples positives (*cf.* [FF1], chap. 10, Lemme 5.4.1). Pour chaque n, on peut écrire $g_n = f_n \circ Y$, où f_n est une variable aléatoire simple positive, nulle sur $Y(\Omega)^c$; de plus, la suite (f_n) est croissante. Alors $f := \sup_n f_n$ est une variable aléatoire positive et $g = f \circ Y$. Si, enfin, g est quelconque, on détermine deux variables aléatoires positives f_1 et f_2 telles que $g^+ = f_1 \circ Y$ et $g^- = f_2 \circ Y$ et on a : $g = (f_1 - f_2) \circ Y$. □

Si (Y_i) $(i \in I)$ est une famille de variables aléatoires réelles définies sur le même espace probabilisé $(\Omega, \mathfrak{A}, \mathrm{P})$, on définit la *tribu engendrée par cette famille* comme étant la tribu, notée $\mathfrak{T}((Y_i)\ (i \in I))$, qui est engendrée par la classe des évènements de la forme $\{Y_i \in B\}$, où $i \in I$ et $B \in \mathcal{B}^1$.

Lorsque l'ensemble I des indices est *fini* et que les variables Y_i prennent leurs valeurs dans $\mathbb{N}$, les éléments de la tribu engendrée par la famille ont une forme explicite, comme établi dans la proposition suivante.

Proposition 1.12.2. *Soient $I = \{i_1 < \dots < i_k\}$ un sous-ensemble fini de $\mathbb{N}$ et (Y_i) $(i \in I)$ une famille de variables aléatoires à valeurs entières, indicée par I. Soit $\mathcal{C}$ la classe des évènements de la forme*

$$\{Y_{i_1} = b_1, \dots, Y_{i_k} = b_k\},$$

où les b_i sont des entiers. Alors, les éléments de la tribu $\mathfrak{T}((Y_i)\ (i \in I))$, engendrée par cette famille, sont aussi les réunions dénombrables d'éléments de $\mathcal{C}$.

Démonstration. En effet, la tribu $\mathfrak{T}((Y_i)\ (i \in I))$ est engendrée par les évènements $\{Y_i = b\}$, où $i \in I$ et $b \in \mathbb{N}$, donc aussi par les éléments de $\mathcal{C}$. Soit $\overline{\mathcal{C}}$ l'ensemble des réunions dénombrables d'éléments de $\mathcal{C}$. Naturellement $\overline{\mathcal{C}} \subset \mathfrak{T}((Y_i)\ (i \in I))$. Il suffit donc de vérifier que $\overline{\mathcal{C}}$ est aussi une tribu. Or, ceci résulte immédiatement du fait que les éléments de $\mathcal{C}$ forment une partition dénombrable de l'ensemble Ω. □

1.13 ESPÉRANCES CONDITIONNELLES

Supposons que $Y = (Y_0, \dots, Y_n)$ est un vecteur aléatoire à $(n+1)$ dimensions et X une variable aléatoire réelle, tous deux définis sur le même espace probabilisé $(\Omega, \mathfrak{A}, \mathrm{P})$. Formons le vecteur aléatoire (X, Y), à $(n+2)$ dimensions et supposons que celui-ci a une *densité de probabilité* $f(x, y)$, où $x \in \mathbb{R}$ et $y \in \mathbb{R}^{n+1}$. On peut alors déterminer la densité du vecteur Y par intégration :

$$f_Y(y) = \int_{\mathbb{R}} f(x, y)\, dx.$$

Considérant une densité arbitraire $g_0(x)$ sur $\mathbb{R}$, on définit la *densité conditionnelle de X à $\{Y = y\}$* comme étant la fonction réelle $f_{X\,|\,Y}(\cdot\,|\,y)$, définie pour tout réel x par :

$$f_{X\,|\,Y}(x\,|\,y) := \begin{cases} \dfrac{f(x, y)}{f_Y(y)}, & \text{si } y \text{ est tel que } f_Y(y) > 0\,; \\ g_0(x), & \text{sinon.} \end{cases}$$

Ensuite, si l'intégrale $\int_{\mathbb{R}} x\, f_{X\,|\,Y}(x\,|\,y)\, dx$ est absolument convergente, on pose

$$\mathbb{E}[X\,|\,Y = y] := \int_{\mathbb{R}} x\, f_{X\,|\,Y}(x\,|\,y)\, dx$$

et on appelle cette intégrale l'*espérance mathématique de la variable aléatoire X conditionnelle à $\{Y = y\}$*. Pour tout $y \in \mathbb{R}^{n+1}$, posons :

$$e(y) := \mathbb{E}[X\,|\,Y = y].$$

Soit P_Y la loi de probabilité du vecteur Y. La fonction e, qui apparaît dans le diagramme suivant, est une variable aléatoire définie sur l'espace probabilisé $(\mathbb{R}^{n+1}, \mathcal{B}^{n+1}, \mathrm{P}_Y)$. La fonction composée $e \circ Y$ est alors une variable aléatoire définie sur $(\Omega, \mathfrak{A}, \mathrm{P})$. On la note $\mathbb{E}[X\,|\,Y]$ et on l'appelle l'*espérance mathématique conditionnelle de X en Y*.

$$
\begin{array}{ccc}
(\Omega, \mathfrak{A}, \mathrm{P}) & \xrightarrow{\mathbb{E}[X \mid Y]} & (\mathbb{R}, \mathcal{B}^1) \\
& {\scriptstyle Y} \searrow \quad \nearrow {\scriptstyle e} & \\
& (\mathbb{R}^{n+1}, \mathcal{B}^{n+1}, \mathrm{P}_Y) &
\end{array}
$$

Considérons maintenant une fonction h de $(n+2)$ variables réelles. Pour tout $y \in \mathbb{R}^{n+1}$, on peut considérer l'intégrale $\int_{\mathbb{R}} h(x,y)\, f_{X \mid Y}(x \mid y)\, dx$. Si elle est absolument convergente, on la pose égale à

$$
\mathbb{E}[h \circ (X,Y) \mid Y = y] := \int_{\mathbb{R}} h(x,y)\, f_{X \mid Y}(x \mid y)\, dx
$$

et on l'appelle l'*espérance mathématique de* $h \circ (X,Y)$ *conditionnelle à* $\{Y = y\}$. Notons encore e la fonction $e : y \mapsto \mathbb{E}[h \circ (X,Y) \mid Y = y]$. C'est un vecteur aléatoire défini sur l'espace $(\mathbb{R}^{n+1}, \mathcal{B}^{n+1}, \mathrm{P}_Y)$. Notons pr_2 la projection $\mathrm{pr}_2 : (x,y) \mapsto y$. On définit l'*espérance mathématique conditionnelle de* $h \circ (X,Y)$ *en* Y comme étant la *variable aléatoire*, notée $\mathbb{E}[h \circ (X,Y) \mid Y]$, définie par le diagramme :

$$
\begin{array}{ccc}
(\Omega, \mathfrak{A}, \mathrm{P}) & \xrightarrow{\quad \mathbb{E}[h \circ (X,Y) \mid Y] \quad} & (\mathbb{R}, \mathcal{B}^1) \\
{\scriptstyle (X,Y)} \searrow & & \nearrow {\scriptstyle e} \\
(\mathbb{R}^{n+2}, \mathcal{B}^{n+2}, \mathrm{P}_{(X,Y)}) & \xrightarrow{\mathrm{pr}_2} & (\mathbb{R}^{n+1}, \mathcal{B}^{n+1}, \mathrm{P}_Y)
\end{array}
$$

Pour des raisons typographiques, on pose aussi $\mathbb{E}_Y[X] := \mathbb{E}[X \mid Y]$. Nous rassemblons ci-dessous quelques propriétés importantes de l'espérance conditionnelle. Nous leur donnons le nom de *Règles* pour mettre en évidence leur caractère *opératoire*. Les symboles f, g, h, ... qui y figurent, désignent des fonctions mesurables dont les arguments sont en évidence de par l'écriture.

Règle 1. *Si X et Y sont indépendantes, alors $\mathbb{E}[g \circ X \mid Y] = \mathbb{E}[g \circ X]$. En particulier, si $g \circ X := X$, on obtient $\mathbb{E}[X \mid Y] = \mathbb{E}[X]$.*

Règle 2. *On a : $\mathbb{E}[h \circ Y \mid Y] = h \circ Y$. En particulier, en prenant $h \circ Y = \mathbb{E}[X \mid Y]$, on obtient : $\mathbb{E}_Y[\mathbb{E}_Y[X]] = \mathbb{E}_Y[X]$, ce qui signifie que l'opérateur $\mathbb{E}_Y$ est idempotent.*

Règle 3. *L'opérateur $\mathbb{E}_Y$ est linéaire : $\mathbb{E}_Y[X_1 + X_2] = \mathbb{E}_Y[X_1] + \mathbb{E}_Y[X_2]$ et $\mathbb{E}_Y[c\,X] = c\,\mathbb{E}_Y[X]$.*

Règle 4. *On a : $\mathbb{E}[(g \circ X)(h \circ Y) \mid Y] = (h \circ Y)\,\mathbb{E}[g \circ X \mid Y]$. Autrement dit, dans le calcul de l'espérance mathématique conditionnelle en Y, une fonction de Y seul se comporte comme une constante.*

Règle 5. *On a : $\mathbb{E}[\mathbb{E}[u \circ (X,Y) \mid Y]] = \mathbb{E}[u \circ (X,Y)]$. Autrement dit, l'espérance de l'espérance conditionnelle est égale à l'espérance.*

Notons trois cas particuliers. En prenant $u(x,y) := g(x)h(y)$, on obtient : $\mathbb{E}[(g\circ X)(h\circ Y)] = \mathbb{E}[\mathbb{E}[(g\circ X)(h\circ Y)\,|\,Y]]$, qui est égal à $\mathbb{E}[(h\circ Y)\,\mathbb{E}[g\circ X\,|\,Y]]$, d'après la Règle 4. On a ainsi :

$$\mathbb{E}[(h\circ Y)\,\mathbb{E}[(g\circ X)\,|\,Y]] = \mathbb{E}[(g\circ X)(h\circ Y)]. \tag{1.13.1}$$

Prenons maintenant $u(x,y) := g(x)$. Alors

$$\mathbb{E}[\mathbb{E}[(g\circ X)\,|\,Y]] = \mathbb{E}[g\circ X]. \tag{1.13.2}$$

Prenons, enfin, $u(x,y) := g(x) = I_{]-\infty,t]}(x)$, de sorte que $\mathbb{E}[g\circ X] = \int g(x)\,d\mathrm{P}_X(x) = \mathrm{P}\{X\le t\}$. Par ailleurs,

$$\begin{aligned}\mathbb{E}[\mathbb{E}[g\circ X\,|\,Y]] &= \int_{\mathbb{R}^{n+1}} \mathbb{E}[g\circ X\,|\,Y=y]\,d\mathrm{P}_Y(y)\\ &= \int_{\mathbb{R}^{n+1}} \mathbb{E}[I_{\{X\le t\}}\,|\,Y=y]\,d\mathrm{P}_Y(y)\\ &= \int_{\mathbb{R}^{n+1}} \mathrm{P}\{X\le t\,|\,Y=y\}\,d\mathrm{P}_Y(y).\end{aligned}$$

On a donc la formule :

$$\mathrm{P}\{X\le t\} = \int_{\mathbb{R}^{n+1}} \mathrm{P}\{X\le t\,|\,Y=y\}\,d\mathrm{P}_Y(y). \tag{1.13.3}$$

Toutes ces formules sont démontrées dans [FF1], chap. 12, § 4, dans le cas où les variables sont absolument continues, la variable Y pouvant être à une ou plusieurs dimensions et s'appliquent, presque *ne varietur*, dans le cas où Y est à *plusieurs* dimensions. Nous nous dispensons donc de faire de nouvelles démonstrations. En revanche, nous établissons l'extension de la formule (1.13.2), au cas où l'opérateur $\mathbb{E}$ est remplacé par l'opérateur $\mathbb{E}_Z$, où Z est une nouvelle variable aléatoire. Dans cette extension, on suppose connue la loi du triplet (X,Y,Z).

Règle 6. *On a*

$$\mathbb{E}_Z[\mathbb{E}_Z[(g\circ X)\,|\,Y]] = \mathbb{E}_Z[g\circ X], \tag{1.13.4}$$

une identité qui doit se comprendre comme

$$\mathbb{E}_Z[\mathbb{E}_{(Y,Z)}[(g\circ X)]] = \mathbb{E}_Z[g\circ X], \tag{1.13.5}$$

ou bien encore comme

$$\mathbb{E}[\mathbb{E}[(g\circ X)\,|\,(Y,Z)]\,|\,Z] = \mathbb{E}[(g\circ X)\,|\,Z]. \tag{1.13.6}$$

Démonstration. Faisons la démonstration dans le cas où les variables sont absolument continues. Par définition de l'espérance conditionnelle, on a : $\mathbb{E}[(g\circ X)\,|\,(Y,Z)] = e_1\circ(Y,Z)$, avec $e_1(y,z) = \int g(x)\,f_{X\,|\,(Y,Z)}(x\,|\,(y,z))\,dx$. Puis, $\mathbb{E}[\mathbb{E}[(g\circ X)\,|\,(Y,Z)]\,Z] = \mathbb{E}[e_1\circ(Y,Z)\,|\,Z] = e_2\circ Z$, avec $e_2(z) = \int e_1(y,z)\,f_{Y\,|\,Z}(y\,|\,z)\,dy = \iint g(x)\,f_{X\,|\,(Y,Z)}(x\,|\,(y,z))\,f_{Y\,|\,Z}(y\,|\,z)\,dx\,dy$.
Cette dernière intégrale vaut :

$$\iint g(x)\,\frac{f_{(X,Y,Z)}(x,y,z)}{f_{(Y,Z)}(y,z)}\,\frac{f_{(Y,Z)}(y,z)}{f_Z(z)}\,dx\,dy$$

$$= \int \frac{g(x)}{f_Z(z)} \Big(\int f_{(X,Y,Z)}(x,y,z)\, dy \Big) dx = \int \frac{g(x)}{f_Z(z)} f_{(X,Z)}(x,z)\, dx$$
$$= \int g(x)\, \frac{f_{(X,Z)}(x,z)}{f_Z(z)}\, dx = \int g(x)\, f_{X\,|\,Z}(x\,|\,z)\, dx.$$

D'autre part, $\mathbb{E}[(g \circ X)\,|\,Z] = e_3 \circ Z$ avec $e_3(z) = \int g(x)\, f_{X\,|\,Z}(x\,|\,z)\, dx$. On a donc $e_2(z) = e_3(z)$ et par conséquent l'identité (1.13.6) est vérifiée. □

Reprenons la formule donnant l'espérance mathématique de $h \circ (X,Y)$ conditionnelle à $\{Y = y\}$, en supposant que X et Y sont des *vecteurs* aléatoires, disons à valeurs dans $\mathbb{R}^p$ et $\mathbb{R}^q$. Naturellement, la fonction h reste une fonction *réelle*. La densité $f_{X\,|\,Y}(x\,|\,y)$ se définit de façon analogue et l'on a $f_{X\,|\,Y}(x\,|\,y) = f_X(x)$ en tous les y où $f_Y(y) > 0$, lorsque X et Y sont *indépendants*. Supposons cette dernière condition satisfaite et aussi $f_Y(y) > 0$ en tout y. Alors

$$\mathbb{E}[h \circ (X,Y)\,|\,Y = y] = \int_{\mathbb{R}^p} h(x,y)\, f_{X\,|\,Y}(x\,|\,y)\, dx = \int_{\mathbb{R}^p} h(x,y)\, f_X(x)\, dx,$$

soit

$$\mathbb{E}[h \circ (X,Y)\,|\,Y = y] = \mathbb{E}[h(X,y)]. \tag{1.13.7}$$

Prenons $p = q$ et soient $B_1, \ldots, B_p$ des boréliens de la droite. Notons $X_1, \ldots, X_p$ et $Y_1, \ldots, Y_p$ les composantes des vecteurs aléatoires X et Y. Alors

$$\mathrm{P}\{\bigcap_{i=1}^{p} \{(X_i + Y_i) \in B_i\}\,|\,Y = y\} = \mathbb{E}[I_{\{\bigcap_{i=1}^{p}\{(X_i+Y_i)\in B_i\}\}}\,|\,Y = y]$$
$$= \mathbb{E}[\prod_{i=1}^{p} I_{\{(X_i+Y_i)\in B_i\}}\,|\,Y = y] = \mathbb{E}[\prod_{i=1}^{p} I_{B_i} \circ (X_i + Y_i)\,|\,Y = y]$$
$$= \mathbb{E}[h \circ (X,Y)\,|\,Y = y] = \mathbb{E}[h \circ (X,y)],$$

en prenant pour h la fonction $h(x,y) := \prod_{i=1}^{p} I_{B_i}(x_i + y_i)$. Si donc X et Y sont *indépendants*, on a la formule :

$$\mathrm{P}\{\bigcap_{i=1}^{p} \{(X_i + Y_i) \in B_i\}\,|\,Y = y\} = \mathrm{P}\{\bigcap_{i=1}^{p} \{(X_i + y_i) \in B_i\}\}. \tag{1.13.8}$$

Ainsi, la loi conditionnelle de $X + Y$, conditionnelle à $\{Y = y\}$, se déduit de celle de X par une simple translation.

Remarque. Toutes ces formules, qui ont été démontrées dans le cas absolument continu, sont encore valables dans le cas général. Notons, enfin, que les formules qui comportent des espérances *conditionnelles* ne sont vraies que *presque sûrement*.

Conservons les mêmes notations pour X et Y. Nous donnons un critère pour qu'une variable aléatoire réelle Z soit égale, presque sûrement, à l'espérance conditionnnelle $\mathbb{E}[X\,|\,Y]$.

Proposition 1.13.1. *Soient X et Z deux variables aléatoires réelles et Y une variable aléatoire à $(n+1)$ dimensions, toutes définies sur un même espace probabilisé $(\Omega, \mathfrak{A}, \mathrm{P})$. Supposons que Z soit de plus mesurable par rapport à $(\Omega, \mathfrak{T}(Y))$. Alors $Z = \mathbb{E}[X \mid Y]$ presque sûrement, si et seulement si pour tout évènement $A \in \mathfrak{T}(Y)$ on a : $\mathbb{E}[I_A Z] = \mathbb{E}[I_A X]$.*

Démonstration. En effet, si cette dernière condition est vérifiée, on a : $\mathbb{E}[I_A Z] = \mathbb{E}[I_A X] = \mathbb{E}[\,\mathbb{E}[I_A \mathbb{E}[X \mid Y]\,]\,] = \mathbb{E}[I_A \mathbb{E}[X|Y]\,]$, d'où

$$\mathbb{E}[I_A(Z - \mathbb{E}[X|Y])\,] = 0.$$

Comme Z et $\mathbb{E}[X \mid Y]$ sont toutes deux des variables aléatoires par rapport à $(\Omega, \mathfrak{T}(Y))$, on peut prendre $A := \{Z - \mathbb{E}[X \mid Y] \geq 0\}$. Il en résulte que $\mathbb{E}[(I_A(Z - \mathbb{E}[X|Y])^+\,] = 0$, ce qui implique $(Z - \mathbb{E}[X \mid Y])^+ = 0$ p.s. De la même façon, $(Z - \mathbb{E}[X \mid Y])^- = 0$ p.s. et donc $Z = \mathbb{E}[X \mid Y]$ p.s.

Réciproquement, cette dernière propriété entraîne que, pour tout $A \in \mathfrak{T}(Y)$, on a : $\mathbb{E}[I_A Z] = \mathbb{E}[I_A \mathbb{E}[X \mid Y]\,] = \mathbb{E}[\,\mathbb{E}[I_A X \mid Y]\,] = \mathbb{E}[I_A X]$. □

Chapitre 2

Temps d'arrêt

La notion de *temps d'arrêt* intervient de façon essentielle dans l'étude des processus stochastiques. Nous donnons ici les principaux résultats les concernant, du moins pour les temps d'arrêt à *valeurs entières*.

2.1 GÉNÉRALITÉS SUR LES PROCESSUS STOCHASTIQUES

Sous le nom de *processus stochastique*, on entend un modèle permettant d'étudier un phénomène aléatoire évoluant au cours du temps. Pour le décrire, on suppose donnés :

(1) un espace probabilisé $(\Omega, \mathfrak{A}, \mathrm{P})$;

(2) un espace mesurable $(E, \mathcal{B})$, où E est appelé *ensemble des états* (du processus) ;

(3) une famille $(Y_t)_{t \in T}$ de variables aléatoires définies sur $(\Omega, \mathfrak{A}, \mathrm{P})$ à valeurs dans E.

L'ensemble T des indices est l'espace des *temps*. Pour $\omega \in \Omega$ et $t \in T$ la quantité $Y_t(\omega)$ est appelée *état du processus à l'instant* t. Étant donné $\omega \in \Omega$, l'application qui à tout t dans T fait correspondre $Y_t(\omega)$, l'état du processus à l'instant t, est appelée la *trajectoire* de ω. Lorsque E et T sont discrets, toute trajectoire est constituée par une suite de points dans le plan, comme représenté dans la Figure 2.1.

On distingue plusieurs types de processus suivant que T et E sont discrets ou continus. Si $T \subset \mathbb{N}$, on dit que le processus est à *temps discret*. Si $T = [0, 1]$

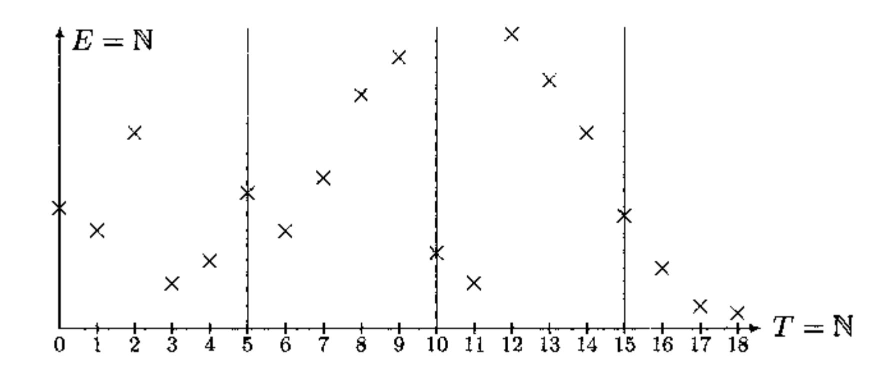


FIGURE 2.1

ou $\mathbb{R}$, on dit que le processus est à *temps continu*. Suivant que E est fini, dénombrable ou continu, on dit que le processus a un *nombre fini d'états*, a un *espace d'états dénombrable*, a un *espace d'états continu*.

Lorsque $T = \mathbb{N}$ (temps discret), le processus stochastique est décrit par une *suite* $Y_0, Y_1, Y_2, \dots$ de variables aléatoires à valeurs dans $(E, \mathcal{B})$. Si E est fini ou dénombrable, ces variables aléatoires sont forcément des variables aléatoires discrètes.

Considérons un processus à temps discret, ou encore une suite de variables aléatoires (Y_n) $(n \geq 0)$. Pour matérialiser la connaissance du processus jusqu'à un instant n, on introduit la tribu des *évènements se produisant jusqu'à l'instant n (inclus)*, comme étant la tribu engendrée par le vecteur $(Y_0, \dots, Y_n)$. On la note $\mathfrak{A}_{0,n}$ ou, plus simplement, $\mathfrak{A}_n$. Elle est donc engendrée par les évènements de la forme $\{Y_i \in B\}$, où $B \in \mathcal{B}^1$ et où $i \in \{0, 1, \dots, n\}$ (*cf.* § 1.12).

On pose $\mathfrak{A}_{-1} = \{\emptyset, \Omega\}$. La suite $(\mathfrak{A}_n)$ de ces tribus est monotone croissante :

$$\mathfrak{A}_{-1} \subset \mathfrak{A}_0 \subset \mathfrak{A}_1 \subset \mathfrak{A}_2 \subset \cdots \tag{2.1.1}$$

Proposition 2.1.1. *Si les variables aléatoires Y_n sont à valeurs entières, ou encore si l'espace des états E est $\mathbb{N}$, chaque évènement $A \in \mathfrak{A}_n$, est réunion dénombrable d'évènements de la forme*

$$\{Y_0 = b_0, Y_1 = b_1, \dots, Y_n = b_n\}$$

où les b_i sont des éléments de E.

Démonstration. Ceci résulte immédiatement de la Proposition 1.12.2 et de la définition même de la tribu $\mathfrak{A}_n$. □

Si m, n sont deux entiers tels que $m \leq n$, on désigne par $\mathfrak{A}_{m,n}$ la tribu engendrée par le vecteur $(Y_m, \dots, Y_n)$ et par $\mathfrak{A}_{n,\infty}$ la tribu engendrée par la famille (Y_i) $(i \geq n)$.

2.2 NOTION DE TEMPS D'ARRÊT

Considérons une suite de parties indépendantes de pile (« p ») ou face (« f »), la probabilité d'apparition de « p » étant $\frac{1}{2}$. Convenons qu'un joueur, engagé dans cette suite de parties, gagne ou perd 1 euro selon que pile ou face apparaît et désignons par Y_n le gain de ce joueur à la $n^{\text{ième}}$ partie. La suite (Y_n) $(n \geq 1)$ de ces gains est une suite de variables aléatoires, indépendantes et identiquement distribuées, de loi commune $\frac{1}{2}(\varepsilon_{-1} + \varepsilon_{+1})$. Posons $X_n := Y_1 + \cdots + Y_n$, une valeur qui n'est autre que le gain total du joueur après la $n^{\text{ième}}$ partie. Pour tout $n \geq 1$, on a $\mathbb{E}[X_n] = 0$, ce qui s'exprime en disant que le jeu est *équitable*.

Or, le joueur n'est pas obligé de jouer indéfiniment, ni même de s'arrêter à un instant donné *fixé à l'avance*. Il peut, au contraire, adopter une *stratégie d'arrêt* et décider d'arrêter le jeu à un instant qu'il juge favorable, par exemple à l'*instant où il mène pour la première fois*. Désignons par T l'instant où le joueur décide de s'arrêter. Cet instant T est une variable aléatoire à valeurs dans $\mathbb{N} \cup \{+\infty\}$, définie sur le même espace probabilisé que la suite (Y_n) $(n \geq 1)$. Si T est presque sûrement à valeurs finies, le gain du joueur à la date T peut alors s'écrire :

$$X_T = Y_1 + \cdots + Y_T.$$

La question à se poser est la suivante : a-t-on encore $\mathbb{E}[X_T] = 0$? Donnons déjà quelques réponses partielles :

(1) si $T = n$ fixé, on a bien $\mathbb{E}[X_n] = 0$.

(2) si T est une variable aléatoire à valeurs dans $\{1, 2, \dots\}$, indépendante de la suite (Y_n) $(n \geq 1)$ et si $\mathbb{E}[T] < +\infty$, l'identité de Wald (*cf.*, par exemple, [FF1], chap. 9, Exercice 6) permet d'écrire :

$$\mathbb{E}[X_T] = \mathbb{E}[Y_1 + \cdots + Y_T] = \mathbb{E}[T]\,\mathbb{E}[Y_1] = 0.$$

Que peut-on dire pour d'autres instants aléatoires T ? Nous allons voir qu'en imposant à T des conditions relativement naturelles et simples, cette propriété de jeu équitable est conservée. Nous sommes ainsi conduits à la notion de *temps d'arrêt*.

Définition 2.2.1. Soit (Y_n) $(n \geq 0)$ une suite de variables aléatoires, définies sur un même espace probabilisé $(\Omega, \mathfrak{A}, \mathrm{P})$ (donc un processus stochastique à temps discret). Soit T une variable aléatoire définie sur le même espace. On dit que T est un *temps d'arrêt* (ou de Markov), *adapté à la suite* (Y_n) $(n \geq 0)$ (on dit encore : un *temps d'arrêt du processus* (Y_n)), si

(a) la variable aléatoire T prend ses valeurs dans $\mathbb{N} \cup \{+\infty\}$;

(b) pour tout $n \geq 0$ l'évènement $\{T = n\}$ appartient à la tribu $\mathfrak{A}_n = \mathfrak{T}(Y_0, \dots, Y_n)$.

La condition (b) postule qu'à la date n, on ne peut décider de la réalisation ou de la non-réalisation de l'évènement $\{T = n\}$ qu'à partir de l'information fournie par le vecteur $(Y_0, \dots, Y_n)$. En d'autres termes, l'évènement $\{T = n\}$ ne dépend pas de $(Y_{n+1}, Y_{n+2}, \dots)$, c'est-à-dire de l'histoire du processus après

la date n. Dans l'exemple donné précédemment, cette condition empêche toute prévision de la part du joueur (pas de délit d'initié !).

Un énoncé équivalent à celui de la condition (b) est de dire que pour tout $n \geq 0$ l'indicatrice $I_{\{T=n\}}$ de l'évènement $\{T = n\}$ est une fonction mesurable du vecteur $(Y_0, \ldots, Y_n)$ (*cf.* Proposition 1.12.1).

Remarque. Dans la définition d'un temps d'arrêt, la condition (b) n'implique pas nécessairement que l'on ait $\mathbb{E}[X_T] = 0$. Reprenons le même exemple (Y_n) $(n \geq 1)$ d'une suite de parties de pile ou face et définissons

$$T := \min\{n : X_n = Y_1 + \cdots + Y_n = 1\}.$$

D'abord, T est bien un temps d'arrêt adapté à (Y_n), car pour tout $n \geq 1$ on a :

$$\{T = n\} = \{Y_1 < 1, Y_1 + Y_2 < 1, \ldots, Y_1 + \cdots + Y_{n-1} < 1, Y_1 + \cdots + Y_n = 1\};$$

ce qui montre que l'évènement $\{T = n\}$ appartient à la tribu $\mathfrak{T}(Y_0, \ldots, Y_n)$. Ensuite, puisque $X_T = 1$, on a $\mathbb{E}[X_T] = 1 > 0$. Si l'on avait $\mathbb{E}[T] < +\infty$, alors, d'après l'identité de Wald généralisée (*cf.* § 2.5), on aurait $\mathbb{E}[X_T] = \mathbb{E}[T]\,\mathbb{E}[Y_1] = 0$, d'où une contradiction. Il en résulte que $\mathbb{E}[T] = +\infty$. Par conséquent, si l'on adoptait cette stratégie pour arrêter le jeu, on serait amené, en moyenne, à jouer des parties fort longues, pouvant occasionner des découverts importants !

Donnons quelques propriétés des temps d'arrêt.

Proposition 2.2.1. *On peut remplacer la condition* (b) *de la définition d'un temps d'arrêt par l'une des conditions suivantes :*
(b′) *pour tout $n \geq 0$ l'évènement $\{T \leq n\}$ appartient à la tribu $\mathfrak{A}_n$;*
(b″) *pour tout $n \geq 0$ l'évènement $\{T > n\}$ appartient à la tribu $\mathfrak{A}_n$.*

Démonstration. Il suffit de démontrer (b′). D'une part, $\{T \leq n\} = \{T = 0\} + \cdots + \{T = n\}$, d'autre part, $\{T = n\} = \{T \leq n\} \setminus \{T \leq n-1\}$; ce qui permet de conclure puisque la suite des tribus $\mathfrak{A}_n$ est croissante, d'après (2.2.1). □

Proposition 2.2.2. *Si T est un temps d'arrêt du processus (Y_n), alors les évènements $\{T < n\}$ et $\{T \geq n\}$ appartiennent à $\mathfrak{A}_{n-1}$.*

Démonstration. Ce résultat découle immédiatement de la Proposition 2.2.1. □

Remarque. Si, pour tout $n \geq 0$, l'évènement $\{T < n\}$ (ou l'évènement $\{T \geq n\}$) appartient à $\mathfrak{A}_n$, alors T n'est pas nécessairement un temps d'arrêt. On peut voir, tout d'abord, que l'argument utilisé dans la démonstration de la Proposition 2.2.1 échoue, puisque l'identité $\{T = n\} = \{T < n+1\} \setminus \{T < n\}$ permet simplement de conclure que $\{T = n\}$ est dans $\mathfrak{A}_{n+1}$. Pour établir pleinement cette remarque, nous fournissons un contre-exemple issu de la théorie des processus de Poisson (*cf.* chap. 3 et 4).

Contre-exemple. Dans un processus de Poisson, on a une suite croissante d'instants (S_n) $(n \geq 1)$ où se produit un évènement privilégié, qu'on appelle « top » (empruntant la terminologie des émissions radio-actives). On désigne

par $N(t)$ le nombre de tops se produisant dans l'intervalle $[0,t]$. On peut interpréter $N(t)$ comme l'indice du *dernier* top se produisant dans $[0,t]$. En posant $S_0 = 0$ on peut écrire :

$$\{N(t) = n\} = \{S_n \leq t, S_{n+1} > t\}.$$

On voit ainsi que pour décider de la réalisation de l'évènement $\{N(t) = n\}$, il faut connaître les valeurs de S_n *et de* S_{n+1}. Par conséquent $N(t)$ n'est pas un temps d'arrêt adapté aux S_n. Pourtant, l'évènement $\{N(t) \geq n\}$ appartient à $\mathfrak{T}(S_1, \ldots, S_n)$, puisque $\{N(t) \geq n\} = \{S_n \leq t\}$. Il en est de même du complémentaire $\{N(t) < n\}$. En revanche, $N(t)+1$ est un temps d'arrêt adapté aux S_n, puisque $\{N(t) + 1 = n\} = \{N(t) = n-1\} = \{S_{n-1} < t, S_n > t\}$. On peut interpréter $N(t) + 1$ comme l'indice du *premier* top se produisant dans $]t, +\infty[$.

Ce contre-exemple montre, en particulier, que si T est un temps d'arrêt, la variable aléatoire $(T-1)$ n'en est pas nécessairement un. En revanche, la variable $(T+k)$, avec $k \geq 0$ en est un. (Voir Proposition 2.2.3 ci-après.)

Proposition 2.2.3. *La variable aléatoire constante $T = k$ (k entier positif) est un temps d'arrêt. Si T est un temps d'arrêt et k un entier positif, alors $T + k$ est aussi un temps d'arrêt.*

Démonstration. Posons $T' := k$. Alors $\{T' = n\}$ est l'évènement certain ou l'évènement impossible, suivant que $n = k$ ou $n \neq k$, deux évènements qui sont trivialement dans $\mathfrak{A}_n$.

Par ailleurs, $\{T + k = n\} = \{T = n - k\}$, qui est l'évènement impossible si $n < k$ et qui appartient à $\mathfrak{A}_{n-k}$ et donc aussi à $\mathfrak{A}_n$, si $n \geq k$. $\square$

2.3 AUTRES PROPRIÉTÉS DES TEMPS D'ARRÊT

L'algèbre des temps d'arrêt est peu riche. Notons toutefois les propriétés suivantes.

Proposition 2.3.1. *Soient S et T deux temps d'arrêt du processus (Y_n). Alors, $S+T$, $S \wedge T$ et $S \vee T$ sont aussi des temps d'arrêt. En particulier, $T \wedge k$ est un temps d'arrêt borné.*

Démonstration. En effet, pour tout $n \geq 0$, on a $\{S + T = n\} = \bigcup_{0 \leq k \leq n} \{S = k\}\{T = n - k\}$, qui appartient à $\mathfrak{A}_n$, puisque chacun des termes y appartient. De même, $\{S \wedge T > n\} = \{S > n\} \cap \{T > n\}$ et $\{S \vee T \leq n\} = \{S \leq n\} \cap \{T \leq n\}$. On peut alors conclure à l'aide de la Proposition 2.2.1. $\square$

Remarque. Soient S, T deux temps d'arrêt tels que $S < T$. Alors $U := T - S$ n'est *pas* nécessairement un temps d'arrêt.

Pour le voir, reprenons le contre-exemple du processus de Poisson $(N(t))$ donné dans le paragraphe précédent, avec la suite croissante (S_n) des apparitions des tops. D'abord $S := 1$ est un temps d'arrêt, également $T := N(t) + 1$, mais, comme on a vu, $U := T - S = N(t)$ n'est pas un temps d'arrêt.

Revenons au cas général d'un processus (Y_n) $(n \geq 0)$ et d'un temps d'arrêt T de ce processus, que nous supposons à valeurs *finies*. On a défini la tribu $\mathfrak{A}_n = \mathfrak{T}(Y_0, \ldots, Y_n)$. Comment définir la tribu $\mathfrak{A}_T$ des évènemens *survenant jusqu'au temps d'arrêt T (inclus)*?

Proposition 2.3.2. *L'ensemble noté $\mathfrak{A}_T$ des évènements $A \in \mathfrak{A}$ ayant la propriété que, pour tout $n \geq 0$, l'intersection $A \cap \{T = n\}$ appartient à $\mathfrak{A}_n$ est une tribu.*

Démonstration. D'abord, $\Omega \cap \{T = n\} = \{T = n\}$, qui appartient à $\mathfrak{A}_n$, puisque T est un temps d'arrêt, d'où $\Omega \in \mathfrak{A}_T$. Ensuite, soit (A_m) une suite d'éléments de $\mathfrak{A}_T$; comme $\bigcup_m A_m \cap \{T = n\} = \bigcup_m (A_m \cap \{T = n\})$, on voit que la réunion de ces évènements appartient à $\mathfrak{A}_n$, puisque chaque terme y appartient. Enfin, considérons $A \in \mathfrak{A}_T$. Alors $A^c \cap \{T = n\}$ est la différence propre $\{T = n\} \setminus (A \cap \{T = n\})$ de deux évènements qui appartiennent à $\mathfrak{A}_n$; il appartient donc à $\mathfrak{A}_n$ et par suite $A^c \in \mathfrak{A}_T$. □

Proposition 2.3.3. *Soient S et T deux temps d'arrêt. Alors*

$$S \leq T \Longrightarrow \mathfrak{A}_S \subset \mathfrak{A}_T.$$

Démonstration. Prenons $A \in \mathfrak{A}_S$. Comme $S \leq T$, on peut écrire

$$A \cap \{T = n\} = \bigcup_{0 \leq k \leq n} A \cap \{S = k\} \cap \{T = n\}.$$

Or, pour $0 \leq k \leq n$, l'évènement $A \cap \{S = k\}$ est dans $\mathfrak{A}_k$, donc aussi dans $\mathfrak{A}_n$; la réunion est donc aussi dans $\mathfrak{A}_n$. □

Supposons T à valeurs finies. Pour chaque épreuve ω la quantité $Y_{T(\omega)}(\omega)$ est un certain état (élément de E). L'application Y_T envoyant ω sur $Y_{T(\omega)}(\omega)$ est appelée *état du processus à l'instant T*.

Proposition 2.3.4. *Soit T un temps d'arrêt du processus (Y_n) $(n \geq 0)$. L'application Y_T est une variable aléatoire (par rapport à $(\Omega, \mathfrak{A})$). De plus, T et Y_T sont aussi des variables aléatoires par rapport à $(\Omega, \mathfrak{A}_T)$.*

Démonstration. En effet, si B appartient à la tribu définie sur l'ensemble E des états, on a :

$$\{Y_T \in B\} = \bigcup_{k \geq 0} \{T = k, Y_T \in B\} = \bigcup_{k \geq 0} \{T = k, Y_k \in B\}.$$

Chacun des évènements de cette réunion appartient à la tribu $\mathfrak{A}$, donc aussi la réunion. Pour démontrer que T est une variable aléatoire par rapport à $(\Omega, \mathfrak{A}_T)$, il suffit de démontrer que, pour tout entier $k \geq 0$, on a $\{T = k\} \in \mathfrak{A}_T$, ou encore que, pout tout $n \geq 0$, on a $\{T = k\} \cap \{T = n\} \in \mathfrak{A}_n$. Or $\{T = k\} \cap \{T = n\}$ est vide, si $k \neq n$ ou vaut $\{T = n\}$, si $k = n$. Il est donc bien dans $\mathfrak{A}_n$.

De même, $\{Y_T \in B\}$ appartient à $\mathfrak{A}_T$, si et seulement si, pour tout $n \geq 0$, l'évènement $\{Y_T \in B\} \cap \{T = n\}$ appartient à $\mathfrak{A}_n$. Or, cet évènement peut se récrire $\{Y_n \in B, T = n\}$, qui est bien dans $\mathfrak{A}_n$. □

2.4 PROCESSUS DE VARIABLES INDÉPENDANTES

Illustrons l'utilité de cette notion de temps d'arrêt dans l'étude des processus de variables indépendantes. Considérons un processus $(\Omega, \mathfrak{A}, \mathrm{P}, (Y_n), E)$, où $E = \mathbb{N}$ ou une partie finie de $\mathbb{N}$ et où les variables aléatoires Y_n sont *mutuellement indépendantes.* Pour chaque $n \geq 0$ soit Q_n la loi de Y_n. Pour toute suite $(b_0, \ldots, b_{n+1})$ d'éléments de E, on a donc

$$\mathrm{P}\{Y_0 = b_0, \ldots, Y_{n+1} = b_{n+1}\} = \mathrm{P}\{Y_0 = b_0\} \cdots \mathrm{P}\{Y_{n+1} = b_{n+1}\}. \qquad (2.4.1)$$

Lemme 2.4.1. *Soit A un évènement appartenant à $\mathfrak{A}_n$. Alors pour tout $B_{n+1} \subset E$ on a :*

$$\mathrm{P}\{A, Y_{n+1} \in B_{n+1}\} = \mathrm{P}(A) Q_{n+1}(B_{n+1}) = \mathrm{P}(A)\mathrm{P}\{Y_{n+1} \in B_{n+1}\}.$$

Démonstration. D'après la Proposition 2.1.1, chaque évènement $A \in \mathfrak{A}_n$ est réunion dénombrable d'évènements de la forme $\{Y_1 = b_1, \ldots, Y_n = b_n\}$. On écrit symboliquement $A = \sum_{b_1,\ldots,b_n} \{Y_1 = b_1, \ldots, Y_n = b_n\}$. L'évènement $\{Y_{n+1} \in B_{n+1}\}$ est aussi réunion dénombrable d'évènements de la forme $\{Y_{n+1} = b_{n+1}\}$. D'après (2.4.1), on peut donc écrire, les sommations

$$\begin{aligned}
\mathrm{P}\{A, Y_{n+1} \in B_{n+1}\} &= \sum_{b_1,\ldots,b_n,b_{n+1}} \mathrm{P}\{Y_1 = b_1, \ldots, Y_n = b_n, Y_{n+1} = b_{n+1}\} \\
&= \sum_{b_1,\ldots,b_n,b_{n+1}} \mathrm{P}\{Y_1 = b_1, \ldots, Y_n = b_n\}\mathrm{P}\{Y_{n+1} = b_{n+1}\} \\
&= \sum_{b_1,\ldots,b_n} \mathrm{P}\{Y_1 = b_1, \ldots, Y_n = b_n\} \sum_{b_{n+1}} \mathrm{P}\{Y_{n+1} = b_{n+1}\} \\
&= \mathrm{P}(A)\mathrm{P}\{Y_{n+1} \in B_{n+1}\}. \quad \square
\end{aligned}$$

Dans le théorème suivant, on suppose que la suite (Y_n) $(n \geq 0)$ est formée de variables aléatoires, *indépendantes et de même loi.* Leur loi commune est notée Q. On démontre que le résultat précédent est encore valable, si l'on remplace la suite des instants fixés par une suite strictement croissante de temps d'arrêt.

Théorème 2.4.2 (Théorème des temps d'arrêt de Doob). *Soit $(T_n)_{n\geq 1}$ une suite strictement croissante de temps d'arrêt à valeurs finies, soit A un évènement appartenant à $\mathfrak{A}_{T_1}$ et soient $B_1, \ldots, B_n$ des sous-ensembles de E. Alors on a la relation :*

$$\mathrm{P}\{A, Y_{T_1+1} \in B_1, \ldots, Y_{T_n+1} \in B_n\} = \mathrm{P}(A) Q(B_1) \cdots Q(B_n).$$

Démonstration. En considérant le système complet d'évènements $(\{T_n = k\})$ $(k \geq n-1)$, on peut obtenir la décomposition :

$$\begin{aligned}
&\{A, Y_{T_1+1} \in B_1, \ldots, Y_{T_n+1} \in B_n\} \\
&\qquad = \sum_{k \geq n-1} \{A, Y_{T_1+1} \in B_1, \ldots, Y_{T_n+1} \in B_n, T_n = k\}
\end{aligned}$$

$$= \sum_{k \geq n-1} \{A, Y_{T_1+1} \in B_1, \dots, Y_{T_{n-1}+1} \in B_{n-1}, T_n = k, Y_{k+1} \in B_n\}.$$

Or, du fait que la suite des temps d'arrêt (T_n) est strictement croissante, l'évènement $\{A, Y_{T_1+1} \in B_1, \dots, Y_{T_{n-1}+1} \in B_{n-1}, T_n = k\}$ est la réunion d'évènements de la forme

$$\{A, Y_{i_1+1} \in B_1, \dots, Y_{i_{n-1}+1} \in B_{n-1}, T_1 = i_1, \dots, T_{n-1} = i_{n-1}, T_n = k\},$$

où $0 \leq i_1 < \cdots < i_{n-1} < k$. Comme $A \in \mathfrak{A}_{T_1}$, on a : $A \cap \{T_1 = i_1\} \in \mathfrak{A}_{i_1} \subset \mathfrak{A}_k$. Ensuite, $\{Y_{i_1+1} \in B_1, \dots, Y_{i_{n-1}+1} \in B_{n-1}\} \in \mathfrak{A}_{i_{n-1}+1} \subset \mathfrak{A}_k$. Enfin, $\{T_1 = i_1\} \in \mathfrak{A}_{i_1} \subset \mathfrak{A}_k$, ... , $\{T_{n-1} = i_{n-1}\} \in \mathfrak{A}_{i_{n-1}} \subset \mathfrak{A}_k$, $\{T_n = k\} \in \mathfrak{A}_k$, du fait que les T_i sont des temps d'arrêt. Il en résulte que l'évènement $\{A, Y_{T_1+1} \in B_1, \dots, Y_{T_{n-1}+1} \in B_{n-1}, T_n = k\}$ appartient à $\mathfrak{A}_k$. On peut alors appliquer le Lemme 2.4.1 et écrire :

$$\begin{aligned}
&\mathrm{P}\{A, Y_{T_1+1} \in B_1, \dots, Y_{T_{n-1}+1} \in B_{n-1}, T_n = k, Y_{k+1} \in B_n\} \\
&\quad = \mathrm{P}\{A, Y_{T_1+1} \in B_1, \dots, Y_{T_{n-1}+1} \in B_{n-1}, T_n = k\} \, \mathrm{P}\{Y_{k+1} \in B_n\} \\
&\quad = \mathrm{P}\{A, Y_{T_1+1} \in B_1, \dots, Y_{T_{n-1}+1} \in B_{n-1}, T_n = k\} \, Q(B_n)
\end{aligned}$$

En sommant par rapport à k, on en déduit :

$$\begin{aligned}
&\mathrm{P}\{A, Y_{T_1+1} \in B_1, \dots, Y_{T_{n-1}+1} \in B_{n-1}, Y_{T_n+1} \in B_n\} \\
&\quad = \sum_{k \geq n-1} \mathrm{P}\{A, Y_{T_1+1} \in B_1, \dots, Y_{T_{n-1}+1} \in B_{n-1}, T_n = k\} \, Q(B_n) \\
&\quad = \mathrm{P}\{A, Y_{T_1+1} \in B_1, \dots, Y_{T_{n-1}+1} \in B_{n-1}\} \, Q(B_n).
\end{aligned}$$

D'où, par récurrence sur n,

$$\begin{aligned}
&\mathrm{P}\{A, Y_{T_1+1} \in B_1, \dots, Y_{T_{n-1}+1} \in B_{n-1}, Y_{T_n+1} \in B_n\} \\
&\quad = \mathrm{P}\{A, Y_{T_1+1} \in B_1, \dots, Y_{T_{n-2}+1} \in B_{n-2}\} \, Q(B_{n-1}) \, Q(B_n). \\
&\quad = \cdots \\
&\quad = P(A) Q(B_1) \cdots Q(B_{n-1}) \, Q(B_n). \quad \square
\end{aligned}$$

Conservons les mêmes notations que dans le précédent théorème. Pour chaque $n \geq 1$ et chaque $k \geq 1$, la suite $(T_1, \dots, T_n, T_n + 1, T_n + 2, \dots, T_n + k)$ est encore une suite strictement croissante de temps d'arrêt. Dans l'identité

$$\begin{aligned}
&\mathrm{P}\{A, Y_{T_1+1} \in B_1, \dots, Y_{T_n+1} \in B_n, Y_{T_n+2} \in B_{n+1}, \dots, Y_{T_n+k} \in B_{n+k-1}\} \\
&\qquad\qquad = \mathrm{P}(A) Q(B_1) \cdots Q(B_n) Q(B_{n+1}) \cdots Q(B_{n+k-1}),
\end{aligned}$$

prenons $A = \Omega$ et $B_1 = \cdots = B_{n-1} = E$, on en déduit :

$$\mathrm{P}\{Y_{T_n+1} \in B_n, \dots, Y_{T_n+k} \in B_{n+k-1}\} = Q(B_n) \cdots Q(B_{n+k-1}).$$

Le processus $(Y_{T_1+1}, Y_{T_1+2}, \dots)$ est donc encore un processus de variables indépendantes, de même loi.

2.5 IDENTITÉ DE WALD

Pour terminer ce chapitre, établissons la généralisation suivante de l'identité classique de Wald.

Théorème 2.5.1 (Identité de Wald). *Soient (X_n) $(n \geq 1)$ une suite de variables aléatoires indépendantes, identiquement distribuées, vérifiant $\mathbb{E}[\,|X_1|\,] < +\infty$ et N un temps d'arrêt, à valeurs dans $\{1, 2, \dots\}$, adapté à la suite (X_n) $(n \geq 1)$, d'espérance mathématique finie : $\mathbb{E}[N] < +\infty$. Alors*

$$\mathbb{E}\Big[\sum_{n=1}^{N} X_n\Big] = \mathbb{E}[N]\,\mathbb{E}[X_1].$$

Démonstration. On a d'abord l'identité évidente :

$$\sum_{n=1}^{N} X_n = \sum_{n\geq 1} X_n\, I_{\{n\leq N\}} = X_1 + \sum_{n\geq 2} X_n\, I_{\{n\leq N\}}. \tag{2.5.1}$$

Ensuite, la Proposition 2.2.2 implique que, pour tout $n \geq 2$, l'indicatrice $I_{\{n\leq N\}}$ ne dépend que de $(X_1, \dots, X_{n-1})$; elle est donc indépendante de X_n ; d'où $\mathbb{E}[X_n\, I_{\{n\leq N\}}] = \mathbb{E}[X_n]\,\mathbb{E}[I_{\{n\leq N\}}] = \mathbb{E}[X_1]\,\mathrm{P}\{N \geq n\}$. Prenons alors l'espérance mathématique des deux membres de l'identité (2.5.1). On a formellement :

$$\begin{aligned}\mathbb{E}\Big[\sum_{n=1}^{N} X_n\Big] &= \mathbb{E}\Big[\sum_{n\geq 1} X_n\, I_{\{n\leq N\}}\Big] = \sum_{n\geq 1} \mathbb{E}[X_n\, I_{\{n\leq N\}}]\\ &= \mathbb{E}[X_1] \sum_{n\geq 1} \mathrm{P}\{N \geq n\} = \mathbb{E}[N]\,\mathbb{E}[X_1].\end{aligned}$$

Pour justifier ce calcul formel, il suffit de montrer que la série de terme général $\mathbb{E}[\,|X_n|\, I_{\{n\leq N\}}]$ est convergente. Or, ceci est le cas, puisque :

$$\begin{aligned}\sum_{n\geq 1} \mathbb{E}[\,|X_n|\, I_{\{n\leq N\}}] &= \sum_{n\geq 1} \mathbb{E}[\,|X_n|\,]\,\mathbb{E}[I_{\{n\leq N\}}]\\ &= \mathbb{E}[\,|X_1|\,] \sum_{n\geq 1} \mathrm{P}\{N \geq n\} = \mathbb{E}[\,|X_1|\,]\mathbb{E}[N] < +\infty. \quad \square\end{aligned}$$

2.6 COMPLÉMENTS ET EXERCICES

2.1 Temps d'atteinte
Soit A un sous-ensemble d'états et (Y_n) $(n \geq 0)$ un processus. Le *temps d'atteinte dans l'ensemble A* du processus est défini par :

$$T_A := \inf\{n \geq 0 : Y_n \in A\}.$$

Soit r un entier tel que $r \geq 1$. On définit $T_A^{(r)}$ comme étant le premier instant où le processus (Y_n) atteint le sous-ensemble A pour la $r^{\text{ième}}$ fois. Alors $T_A = T_A^{(1)}$ et $T_A^{(r)}$ $(r \geq 2)$ sont des temps d'arrêt du processus (Y_n).

2.2 Soient (Y_n) $(n \geq 0)$ un processus et A un sous-ensemble de ses états. Le *dernier* instant T'_A, où le processus (Y_n) est dans A, appartient à la tribu engendrée par la suite infinie $(Y_n, Y_{n+1}, \dots)$.

Supposons donné un processus (Z_n) $(n \geq 0)$, de variables aléatoires indépendantes et de même loi, celle-ci étant donnée par $q\varepsilon_{-1} + p\varepsilon_{+1}$ $(p+q=1)$. Pour chaque $n \geq 0$, on pose $Y_n := Z_0 + \cdots + Z_n$ et on note A l'ensemble des entiers négatifs ou nuls. Alors T'_A n'est pas un temps d'arrêt du processus (Y_n) $(n \geq 0)$.

2.3 On considère une suite infinie de parties indépendantes de « pile » ou « face », la probabilité d'obtenir '« pile » étant p et celle d'obtenir « face » étant q $(p+q=1)$. On forme la suite (Y_n) $(n \geq 0)$ des variables aléatoires indicatrices de ces parties. La suite (Y_n) est donc un processus de variables aléatoires indépendantes, de même loi, cette loi étant donnée par $q\varepsilon_0 + p\varepsilon_1$. Se reportant à l'Exercice 2.1, on pose $A := \{1\}$ et on considère la suite des temps d'arrêt $T_A^{(r)}$ $(r \geq 1)$. La loi de $T_A = T_A^{(1)}$ est la loi géométrique et celle de $T_A^{(r)}$ $(r \geq 2)$ est la loi binomiale négative.

2.4 On considère le processus (Y_n) $(n \geq 0)$ décrit dans l'Exercice 2.3 et on définit :

$$T := \begin{cases} 2, & \text{si } Y_0 = 1\,; \\ 3, & \text{si } Y_0 = 0. \end{cases}$$

Alors T est un temps d'arrêt. En revanche, si l'on pose

$$T' := \begin{cases} 2, & \text{si } Y_3 = 1\,; \\ 3, & \text{si } Y_3 = 0\,; \end{cases}$$

alors T' n'est pas un temps d'arrêt.

2.5 On considère, de nouveau, le processus (Z_n) $(n \geq 0)$ décrit dans l'Exercice 2.3, mais on suppose $p = q = \frac{1}{2}$. Pour $n \geq 0$, on pose toujours

$Y_n = Z_0 + Z_1 + \cdots + Z_n$. On désigne par T le temps d'atteinte (voir Exercice 2.2) du processus (Y_n) $(n \geq 0)$ dans l'ensemble $A = \{1\}$.

Établir $\mathrm{P}\{T \leq 2k, Y_{2k} \leq 1\} = \mathrm{P}\{T \leq 2k, Y_{2k} \geq 1\}$, puis $\mathrm{P}\{T \leq 2k\} = 1 - \mathrm{P}\{Y_{2k} = -1\}$. En déduire la fonction de répartition de T.

2.6 Le collectionneur de vignettes
Pierre achète des tablettes de chocolat d'une certaine marque. Dans chaque tablette, il trouve une vignette qu'il peut coller dans un album édité par cette marque. L'album contient m emplacements. Soit $T_1 := 1$, puis T_2 le premier instant après T_1 où il obtient une nouvelle vignette, puis T_3 le premier instant après T_2, où il obtient encore une nouvelle vignette, ... , $T := T_m$ l'instant où son album est achevé. On pose : $U_1 := T_1 = 1$, $U_2 := T_2 - T_1$, $U_3 := T_3 - T_2$, ... , $U_m = T_m - T_{m-1}$.

(a) Déterminer la loi du vecteur $(U_1, U_2, \ldots, U_m)$.

(b) Les variables aléatoires U_1, U_2, ... , U_m sont indépendantes.

(c) En déduire la fonction génératrice de T et son espérance mathématique.

2.7 Le contrôle des naissances
Soit (Y_n) $(n \geq 1)$ une suite de variables aléatoires indépendantes, suivant chacune la loi de Bernoulli $\frac{1}{2}(\varepsilon_0 + \varepsilon_1)$. On convient que, pour tout $n \geq 1$, la variable Y_n est égale à 1, si une famille donnée obtient un garçon lors de la $n^{\text{ième}}$ naissance et 0 dans l'autre cas. Les naissances jumelles ou triples ne sont pas considérées dans ce modèle! Par conséquent, $X_n := Y_1 + \cdots + Y_n$ est le nombre de garçons après les n premières naissances et X_n suit la loi binomiale $B(\frac{1}{2}, n)$. La famille applique un système de contrôle de naissances, qui consiste en la donnée d'un *temps d'arrêt* N par rapport à la suite (Y_n) $(n \geq 1)$. Le nombre de garçons dans une famille pratiquant ce système de contrôle des naissances est alors X_N.

D'après l'identité de Wald généralisée, on a $\mathbb{E}[X_N] = \mathbb{E}[N]\,\mathbb{E}[X_1] = \frac{1}{2}\,\mathbb{E}[N]$. Autrement dit, *quel que soit le système de contrôle de naissances choisi*, la proportion de garçons (et de filles) est la même, en moyenne.

Exemple (a) La famille décide de s'arrêter de procréer dès que, pour la première fois, elle a eu deux garçons consécutivement. Évaluer le nombre moyen d'enfants et de garçons dans la famille.

Exemple (b) La famille décide de s'arrêter de procréer dès que, pour la première fois, un garçon succède immédiatement à une fille. Évaluer le nombre moyen d'enfants et de garçons dans la famille.

2.8 Considérons une suite infinie (Y_n) $(n \geq 0)$ de variables aléatoires indépendantes, suivant chacune la loi de Bernoulli $q\varepsilon_0 + p\varepsilon_1$ avec $0 < p < 1$

et $q = 1 - p$. À l'aide de cette suite, on définit une autre suite (X_n) $(n \geq 0)$ de variables aléatoires, en posant :

$$X_n := \begin{cases} +1, & \text{si l'évènement } \{Y_{2n} = 1, Y_{2n+1} = 0\} \text{ est réalisé;} \\ -1, & \text{si l'évènement } \{Y_{2n} = 0, Y_{2n+1} = 1\} \text{ est réalisé;} \\ 0, & \text{si l'évènement } \{Y_{2n} = Y_{2n+1}\} \text{ est réalisé.} \end{cases}$$

De plus, on pose : $T := \inf\{n \geq 0 : X_n = +1 \text{ ou } X_n = -1\}$.

(a) La variable aléatoire T est un temps d'arrêt adapté à la suite (X_n) $(n \geq 0)$. Il est presque sûrement fini. Déterminer sa loi de probabilité et son espérance mathématique.

(b) Calculer la loi de probabilité de la variable aléatoire X_T.

Chapitre 3

Processus de Poisson

Le processus de Poisson sur la droite est un processus à temps continu et à valeurs entières positives. On dit encore que c'est un *processus de comptage*, que l'on note $\{N(t) : t \geq 0\}$. Il s'agit d'étudier le nombre aléatoire $N(t)$ de certains évènements qui se produisent dans un intervalle de temps $[0, t]$ donné. Sa grande popularité dans les applications vient notamment du fait que beaucoup de calculs le concernant sont explicites.

3.1 DÉFINITION ET PREMIÈRES PROPRIÉTÉS

On se propose d'étudier la répartition dans le temps d'instants aléatoires, appelés *instants-τ*. Dans les applications, ce sont les instants où se produisent certains *évènements* spécifiques, comme, par exemple, les émissions de particules radioactives, les appels dans un central téléphonique, les arrivées de clients devant un guichet, etc. Par analogie avec le premier exemple cité, on appelle ces évènements des *tops*, qui se produisent donc aux dits *instants-τ*.

Définition 3.1.1. Désignons par $N(t)$ $(t \geq 0)$ le nombre de tops se produisant dans l'intervalle de temps $[0, t]$ et supposons que $N(0) = 0$. Le processus $\{N(t) : t \geq 0\}$ est appelé *processus de comptage.*

Tout processus de comptage vérifie les propriétés suivantes :

(a) *Pour tout $t \geq 0$ le nombre $N(t)$ est à valeurs entières positives.*

(b) *La fonction $t \mapsto N(t)$ est croissante.*

(c) *Pour tout couple (a, b) $(0 < a < b)$, la différence $N(b) - N(a)$ représente le nombre de tops se produisant dans l'intervalle de temps $]a, b]$.*

Soient $I_1, \ldots, I_k$ des intervalles sur l'axe des temps, disjoints ou non ; soient $N_1, \ldots, N_k$ le nombre de tops se produisant dans ces intervalles. Le k-uple $(N_1, \ldots, N_k)$ est un point aléatoire à valeurs dans $\mathbb{N}^k$, admettant une loi de probabilité $\mathcal{L}$. Faisons subir à tous les intervalles $I_1, \ldots, I_k$, simultanément, une même translation sur l'axe des temps. Soient $I'_1, \ldots, I'_k$ les intervalles translatés et $N'_1, \ldots, N'_k$ le nombre de tops se produisant dans ces intervalles. Le k-uple $(N'_1, \ldots, N'_k)$ est un point aléatoire à valeurs dans $\mathbb{N}^k$ admettant la loi de probabilité $\mathcal{L}'$.

Définition 3.1.2. Le processus de comptage $\{N(t) : t \geq 0\}$ est dit *à accroissements stationnaires*, si quel que soit k, quelle que soit la suite $(I_1, \ldots, I_k)$ et quelle que soit la translation, les lois $\mathcal{L}$ et $\mathcal{L}'$ coïncident.

Lorsque $k = 2$, on peut reprendre cet énoncé de la façon suivante : si un processus de comptage est à accroissements stationnaires, alors la loi de probabilité du nombre de tops se produisant dans un intervalle de temps donné ne dépend que de la *longueur* de cet intervalle.

Définition 3.1.3. Un processus de comptage est dit *à accroissements indépendants*, si les nombres de tops se produisant dans des intervalles de temps disjoints sont indépendants.

Définition 3.1.4. Un processus de comptage est dit *localement continu en probabilité*, si pour tout $t \geq 0$, on a

$$\lim_{h \downarrow 0} \mathrm{P}\{N(t+h) - N(t) \geq 1\} = 0.$$

Parmi les processus de comptage, ce sont les processus de Poisson, définis ci-après, qui jouent un rôle prépondérant.

Définition 3.1.5 (Processus de Poisson). Un processus de comptage $\{N(t), t \geq 0\}$ est appelé *processus de Poisson*, de *densité* $\lambda > 0$, s'il vérifie les propriétés suivantes :

(i) $N(0) = 0$;

(ii) le processus est à accroissements indépendants ;

(iii) le nombre de tops se produisant dans un intervalle de temps de longueur $t \geq 0$ suit la loi de Poisson de paramètre λt, c'est-à-dire, pour tout $s \geq 0$ et tout $t \geq 0$, on a :

$$\mathrm{P}\{N(s+t) - N(s) = n\} = e^{-\lambda t} \frac{(\lambda t)^n}{n!} \quad (n \geq 0).$$

Voici quelques propriétés que vérifie un processus de Poisson.

Propriété 3.1.1. *Un processus de Poisson est à accroissements stationnaires.*

Démonstration. Faisons la démonstration pour deux intervalles $]a, c]$, $]b, d]$, avec $a < b \leq c < d$. Comme le processus est à accroissements indépendants,

on a :

$$\begin{aligned}
&\mathrm{P}\{N(c)-N(a)=k, N(d)-N(b)=l\}\\
&\qquad=\sum \mathrm{P}\{N(b)-N(a)=i_1, N(c)-N(b)=i_2, N(d)-N(c)=i_3\}\\
&\qquad=\sum \mathrm{P}\{N(b)-N(a)=i_1\}\mathrm{P}\{N(c)-N(b)=i_2\}\mathrm{P}\{N(d)-N(c)=i_3\}\\
&\qquad=\sum e^{-\lambda(d-a)}\frac{(\lambda(b-a))^{i_1}}{i_1!}\frac{(\lambda(c-b))^{i_2}}{i_2!}\frac{(\lambda(d-c))^{i_3}}{i_3!}
\end{aligned}$$

où la somme est sur toutes les suites (i_1, i_2, i_3) d'entiers tels que $i_1+i_2=k$ et $i_2+i_3=l$. Or, cette expression ne dépend que des longueurs des intervalles mis en jeu, non de leurs positions sur la droite. □

Propriété 3.1.2. *Un processus de Poisson est localement continu en probabilité.*

Démonstration. Puisqu'un processus de Poisson est à accroissements stationnaires, on a $\mathrm{P}\{N(t+h)-N(t)\geq 1\}=\mathrm{P}\{N(h)-N(0)\geq 1\}=\mathrm{P}\{N(h)\geq 1\}=1-\mathrm{P}\{N(h)=0\}=1-e^{-\lambda h}\to 0$, lorsque h tend vers 0. □

On voit ainsi qu'un processus de Poisson vérifie les quatre propriétés (1) $N(0)=0$; (2) il est à accroissements indépendants; (3) il est à accroissements stationnaires; (4) il est localement continu en probabilité.

Il est remarquable que ces quatre propriétés, malgré leur caractère *qualitatif*, suffisent à assurer qu'un processus de comptage soit de Poisson. La démonstration de ce fait est donnée dans le paragraphe 6 de ce chapitre.

Propriété 3.1.3. *Soit $\{N(t) : t\geq 0\}$ un processus de Poisson de densité $\lambda>0$. Alors, lorsque h tend vers 0,*

(i) $\mathrm{P}\{N(h)=1\}=\lambda h+o(h)$;

(ii) $\mathrm{P}\{N(h)\geq 1\}=\lambda h+o(h)$.

Démonstration. En effet, $\mathrm{P}\{N(h)=1\}=e^{-\lambda h}\lambda h=\lambda h+o(h)$; puis

$$\begin{aligned}
\mathrm{P}\{N(h)\geq 2\}&=\sum_{k\geq 2} e^{-\lambda h}\frac{(\lambda h)^k}{k!}=e^{-\lambda h}(\lambda h)^2\sum_{k\geq 0}\frac{(\lambda h)^k}{(k+2)!}\\
&\leq e^{-\lambda h}(\lambda h)^2\sum_{k\geq 0}\frac{(\lambda h)^k}{k!}=(\lambda h)^2, \text{ d'où (ii).} \quad □
\end{aligned}$$

Propriété 3.1.4. *Pour tout $t>0$ la probabilité pour qu'il se produise une infinité de tops dans $[0,t]$ est nulle. De façon équivalente : avec probabilité 1, la suite $\tau_1, \tau_2, \ldots$ des instants où se produisent les tops n'admet pas de valeur d'adhérence à distance finie.*

Démonstration. Pour tout $t>0$, on a

$$\mathrm{P}\{N(t)<+\infty\}=\sum_{k\geq 0}\mathrm{P}\{N(t)=k\}=e^{-\lambda t}\sum_{k\geq 0}\frac{(\lambda t)^k}{k!}=1. \quad □$$

Comme $N(t)$ est une variable de Poisson de paramètre λt, on a $\mathbb{E}[N(t)]/t = \lambda$. Ainsi λ est l'espérance mathématique du nombre de tops se produisant par unité de temps. C'est la fréquence moyenne des tops. Comme la variance vaut aussi λt, on en déduit la propriété suivante.

Propriété 3.1.5. *Dans un processus de Poisson $\{N(t) : t \geq 0\}$, de densité λ, on a :*

$$\frac{\sigma(N(t))}{\mathbb{E}[N(t)]} = \frac{1}{\sqrt{\lambda t}},$$

qui tend vers $+\infty$ lors que $t \downarrow 0$.

Par conséquent, pour $t > 0$ petit, les fluctuations autour de la moyenne sont grandes par rapport à la moyenne.

Propriété 3.1.6. *On a $\dfrac{N(t)}{t} \xrightarrow{p.s.} \lambda$, lorsque t tend vers l'infini. Ainsi λ est la densité temporelle d'occurrences des tops.*

Démonstration. Notons $X_1, X_2, \dots, X_{[t]}$ le nombre d'occurrences des tops dans les intervalles $[0,1]$, $]1,2]$, $\dots$, $][t]-1,[t]]$ $(t > 1)$. Les X_k sont des variables aléatoires, indépendantes, identiquement distribuées, d'espérance mathématique λ. D'après la loi forte des grands nombres, on a donc :

$$\frac{N([t])}{[t]} = \frac{X_1 + \cdots + X_{[t]}}{[t]} \xrightarrow{p.s.} \mathbb{E}[X_1] = \lambda\,;$$

et de même,

$$\frac{N(t)}{t} \xrightarrow{p.s.} \lambda,$$

lorsque t tend vers l'infini. □

3.2 TEMPS D'ATTENTE. ESPACEMENTS

Soit $\{N(t) : t \geq 0\}$ un processus de Poisson de densité $\lambda > 0$. On note T_1 le *temps d'attente du premier top* (à partir de $t = 0$) et T_n $(n \geq 2)$ *la durée séparant le $(n-1)^{\text{ième}}$ du $n^{\text{ième}}$ top.* On dit encore que les T_n sont les *espacements.* À la suite (T_n) $(n \geq 1)$, on associe la suite (S_n) $(n \geq 0)$ définie par : $S_0 := 0$ et $S_n := T_1 + \cdots + T_n$ $(n \geq 1)$. Par conséquent, S_n $(n \geq 1)$ est l'instant où se produit le $n^{\text{ième}}$ top ou encore le temps d'attente du $n^{\text{ième}}$ top (à partir de $t = 0$).

Théorème 3.2.1. *La suite (T_n) $(n \geq 1)$ est formée de variables aléatoires indépendantes, dont chacune suit la loi exponentielle de paramètre λ $(\lambda > 0)$.*

Démonstration. Montrons que, pour tout $n \geq 1$, la densité conjointe du vecteur $(T_1, \dots, T_n)$ est le produit $\lambda\, e^{-\lambda t_1} \cdots \lambda\, e^{-\lambda t_n}$ de n densités exponentielles de paramètre λ. Le changement de variables $t_k \mapsto s_k$, où $s_k := t_1 + \cdots + t_k$ $(1 \leq k \leq n)$ montre que la densité conjointe de $(T_1, \dots, T_n)$ au point

$(t_1, \ldots, t_n)$ est égale à la densité conjointe de $(S_1, \ldots, S_n)$ au point $(s_1, \ldots, s_n)$, où $s_1 \leq \cdots \leq s_n$.

Calculons cette dernière densité. Pour ce faire, considérons l'évènement

$$A_n := \{S_1 \in [s_1, s_1 + h_1[, S_2 \in [s_2, s_2 + h_2[, \ldots, S_n \in [s_n, s_n + h_n[\},$$

où les s_i et les h_i ont été choisis de sorte que $0 < s_1 < s_1 + h_1 < s_2 < s_2 + h_2 < \cdots < s_n < s_n + h_n$. Alors $\mathrm{P}(A_n)$ est la probabilité de la conjonction des évènements :

zéro top dans $[0, s_1[$ et exactement un top dans $[s_1, s_1 + h_1[$;
zéro top dans $[s_1 + h_1, s_2[$ et exactement un top dans $[s_2, s_2 + h_2[$;
$\ldots$;
zéro top dans $[s_{n-1} + h_{n-1}, s_n[$ et exactement un top dans $[s_n, s_n + h_n[$.

D'où

$$\begin{aligned}\mathrm{P}(A_n) &= e^{-\lambda s_1} e^{-\lambda h_1} \lambda h_1 e^{-\lambda(s_2 - s_1 - h_1)} e^{-\lambda h_2} \lambda h_2 \\ &\qquad \cdots e^{-\lambda(s_n - s_{n-1} - h_{n-1})} e^{-\lambda h_n} \lambda h_n \\ &= e^{-\lambda s_n} e^{-\lambda h_n} \lambda^n h_1 \cdots h_n,\end{aligned}$$

soit, en divisant par $h_1 \cdots h_n$ et en faisant tendre chaque h_i vers 0,

$$\frac{1}{h_1 \cdots h_n} \mathrm{P}(A_n) \to \lambda^n e^{-\lambda s_n},$$

qui est la densité de $(S_1, \ldots, S_n)$. Comme $s_n = t_1 + \cdots + t_n$, on en déduit que $\lambda e^{-\lambda t_1} \cdots \lambda e^{-\lambda t_n}$ est la densité de $(T_1, \ldots, T_n)$. □

Remarque. Il résulte du précédent théorème que $\mathbb{E}[T_n] = \mathbb{E}[S_n - S_{n-1}] = 1/\lambda$. Ainsi, la densité λ est encore l'inverse de l'espérance mathématique de l'intervalle de temps séparant deux tops consécutifs.

Théorème 3.2.2. *Pour tout $n \geq 1$, la variable $S_n = T_1 + \cdots + T_n$, qui représente l'instant où se produit le $n^{\text{ième}}$ top, suit la loi gamma de paramètre (n, λ), à savoir la loi de densité :*

$$f_{S_n}(s) = \begin{cases} \dfrac{\lambda}{(n-1)!} e^{-\lambda s} (\lambda s)^{n-1}, & \textit{si } s \geq 0\,; \\ 0, & \textit{sinon.} \end{cases}$$

Démonstration. La variable S_n est, en effet, la somme de n variables aléatoires indépendantes, suivant toutes la loi exponentielle de paramètre λ. Elle suit donc une loi gamma de paramètre (n, λ) (*cf.* [FF1], chap. 14, § 8 c)). □

Ce théorème permet de donner une autre définition des processus de Poisson, qui a aussi l'avantage de montrer qu'on peut *construire* effectivement un processus de Poisson à partir d'une suite de variables aléatoires exponentielles. Cette construction fait l'objet du paragraphe suivant.

3.3 LA SECONDE DÉFINITION DES PROCESSUS DE POISSON

Soit donnée une suite (T_n) $(n \geq 1)$ de variables aléatoires indépendantes, dont chacune suit la loi exponentielle de paramètre $\lambda > 0$. On lui associe un processus de comptage $\{N(t) : t \geq 0\}$ en convenant que le $n^{\text{ième}}$ top a lieu à l'instant $S_n := T_1 + \cdots + T_n$. En posant $S_0 := 0$, on suppose donc :

$$N(t) := \sum_{n \geq 1} I_{\{S_n \leq t\}}. \tag{3.3.1}$$

Dans ces conditions, on a le résultat suivant.

Théorème 3.3.1. *Le processus de comptage $\{N(t) : t \geq 0\}$ ainsi défini est un processus de Poisson de densité λ.*

La démonstration résulte des deux propositions suivantes.

Proposition 3.3.2. *Pour tout $t > 0$, la variable aléatoire $N(t)$ représentant le nombre de tops se produisant dans l'intervalle de temps $[0, t]$ suit la loi de Poisson de paramètre λt.*

Démonstration. La démonstration repose sur l'identité :

$$\{S_n \leq t\} = \{N(t) \geq n\} \qquad (t > 0,\ n \geq 0). \tag{3.3.2}$$

En effet, l'instant où se produit le $n^{\text{ième}}$ top est inférieur ou égal à t, si et seulement si, dans l'intervalle de temps $[0, t]$, il s'est produit au moins n tops. On en déduit :

$$\begin{aligned} \mathrm{P}\{N(t) = n\} &= \mathrm{P}\{N(t) \geq n\} - \mathrm{P}\{N(t) \geq n+1\} \\ &= \mathrm{P}\{S_n \leq t\} - \mathrm{P}\{S_{n+1} \leq t\}; \end{aligned}$$

d'où, en vertu du Théorème 3.2.2

$$\begin{aligned} \mathrm{P}\{N(t) = n\} &= \int_0^t \lambda e^{-\lambda x} \frac{(\lambda x)^{n-1}}{(n-1)!}\, dx - \int_0^t \lambda e^{-\lambda x} \frac{(\lambda x)^n}{n!}\, dx \\ &= \int_0^t d\Big(e^{-\lambda x} \frac{(\lambda x)^n}{n!}\Big) = e^{-\lambda t} \frac{(\lambda t)^n}{n!}. \quad \square \end{aligned}$$

Supposons donnés un entier $m \geq 1$, une suite (t_n) $(n \geq 1)$ de nombres réels positifs et une suite (i_n) $(n \geq 1)$ d'entiers positifs. Pour $n \geq 1$, posons $s_n := t_1 + t_2 + \cdots + t_n$ et aussi $s_0 := 0$.

Proposition 3.3.3. *La probabilité que l'on ait i_1 tops dans $[0, s_1]$, i_2 tops dans $]s_1, s_2]$, ... , i_m tops dans $]s_{m-1}, s_m]$ est donnée par :*

$$\begin{aligned} &\mathrm{P}\{N(s_1) = i_1, N(s_2) - N(s_1) = i_2, \ldots, N(s_m) - N(s_{m-1}) = i_m\} \\ &\qquad = \frac{e^{-\lambda t_1}(\lambda t_1)^{i_1}}{i_1!} \frac{e^{-\lambda t_2}(\lambda t_2)^{i_2}}{i_2!} \cdots \frac{e^{-\lambda t_m}(\lambda t_m)^{i_m}}{i_m!}. \end{aligned}$$

Avant de donner la démonstration, établissons le lemme suivant.

Lemme 3.3.4. *Soient i un entier tel que $i \geq 2$ et s, t deux nombres réels tels que $t \geq 0$. Alors*

$$\iint_{s<y\leq x\leq s+t} \lambda^2 \frac{(\lambda(x-y))^{i-2}}{(i-2)!}\, dy\, dx = \frac{(\lambda t)^i}{i!}.$$

Démonstration. Il s'agit du calcul d'une intégrale double dans un triangle :

$$\iint_{s<y\leq x\leq s+t} \lambda^2 \frac{(\lambda(x-y))^{i-2}}{(i-2)!}\, dy\, dx = \int_s^{s+t} dy \int_y^{s+t} \lambda^2 \frac{(\lambda(x-y))^{i-2}}{(i-2)!}\, dx$$

$$= \int_s^{s+t} \lambda^i \frac{(s+t-y)^{i-1}}{(i-1)!}\, dy = \frac{(\lambda t)^i}{i!}. \quad \square$$

Démonstration de la Proposition 3.3.3. La Proposition est d'abord vraie pour $m = 1$. C'est le contenu de la Proposition 3.3.2. Posons :

$$A := \{N(s_1) = i_1, N(s_2) - N(s_1) = i_2, \ldots, N(s_m) - N(s_{m-1}) = i_m\}$$

et définissons les variables suivantes :

$X_1 :=$ plus grand instant-τ dans $[0, s_1]$, si $i_1 \geq 1$;
(non défini si $i_1 = 0$) ;

pour $k = 1, 2, \ldots, m-1$,

$Y_k :=$ plus petit instant-τ dans $]s_k, s_{k+1}]$ si $i_{k+1} \geq 1$;
(non défini si $i_{k+1} = 0$) ;

pour $k = 2, \ldots, m$,

$X_k :=$ plus grand instant-τ dans $]s_{k-1}, s_k]$ si $i_k \geq 2$;
(non défini si $i_k = 0$ ou 1) ;

$Y_m :=$ plus petit instant-τ au-delà de s_m.

Dans l'exemple suivant, où $m = 4$ et $(i_1, i_2, i_3, i_4) = (2, 0, 1, 3)$, les tops ont été représentés par des croix. Les variables Y_1 et X_2 ne sont pas définies, car $i_2 = 0$, la variable X_3 n'est pas définie, car $i_3 = 1$.

0 s_1 s_2 s_3 s_4

X_1 Y_2 Y_3 X_4 Y_4

L'évènement A peut encore s'exprimer comme

$$\{0 \leq X_1 \leq s_1 < Y_1 \leq X_2 \leq s_2 < Y_2 \leq X_3 \leq s_3 < \cdots$$
$$\cdots \leq s_{m-1} < Y_{m-1} \leq X_m \leq s_m < Y_m\},$$

avec la convention suivante :

(a) la relation "$X_1 \leq$ " disparaît de l'évènement A, si $i_1 = 0$;

(b) pour $k = 1, \ldots, m-1$, la relation " $< Y_k$" disparaît de l'évènement A, si $i_{k+1} = 0$;

(c) pour $k = 2, \ldots, m$, la relation "$X_k \leq$ " disparaît de l'évènement A, si $i_k = 0, 1$.

Dans l'exemple traité, l'évènement A s'exprime comme :

$$\{0 \leq X_1 \leq s_1 < s_2 < Y_2 \leq s_3 < Y_3 \leq X_4 \leq s_4 < Y_4\}.$$

Notons $Z_1 < Z_2 < \cdots < Z_l$ la suite des variables X_k, Y_k apparaissant *effectivement* dans A. Par construction-même, $Z_l = Y_m$ et il existe une suite strictement croissante d'entiers $1 \leq b_1 < b_2 < \cdots < b_l$ telle que

$$Z_1 = T_1 + \cdots + T_{b_1}, \quad Z_2 - Z_1 = T_{b_1+1} + \cdots + T_{b_2}, \ \ldots \\ \ldots, Z_l - Z_{l-1} = T_{b_{l-1}+1} + \cdots + T_{b_l}.$$

Posons $a_1 := b_1$, $a_2 := b_2 - b_1$, $\ldots$, $a_l := b_l - b_{l-1}$. Alors $a_l = 1$, puisque $Z_l = Y_m$. Comme les variables aléatoires T_i sont mutuellement indépendantes, par hypothèse, les variables aléatoires Z_1, $Z_2 - Z_1$, $\ldots$, $Z_l - Z_{l-1}$ le sont aussi. De plus, Z_1 suit une loi gamma de paramètre (λ, a_1), puis $Z_2 - Z_1$ une loi gamma de paramètre (λ, a_2), $\ldots$, enfin $Z_l - Z_{l-1}$ une loi gamma de paramètre (λ, a_l).

Le vecteur $(Z_1, Z_2 - Z_1, \ldots, Z_l - Z_{l-1})$ a donc une densité qui est le produit des densités des lois gamma de ses composantes. Par ailleurs, le jacobien de passage du vecteur $(Z_1, Z_2 - Z_1, \ldots, Z_l - Z_{l-1})$ au vecteur $Z := (Z_1, Z_2, \ldots, Z_l)$ est évidemment égal à 1. Par conséquent, la densité de ce dernier vecteur, pour $z_1 \geq 0$, $z_2 \geq 0$, $\ldots$, $z_l \geq 0$, est égale à

$$\frac{\lambda}{(a_1 - 1)!}\, e^{-\lambda z_1} (\lambda z_1)^{a_1 - 1} \frac{\lambda}{(a_2 - 1)!}\, e^{-\lambda(z_2 - z_1)} (\lambda(z_2 - z_1))^{a_2 - 1} \ldots \\ \times \frac{\lambda}{(a_l - 1)!}\, e^{-\lambda(z_l - z_{l-1})} (\lambda(z_l - z_{l-1}))^{a_l - 1},$$

ou encore, en posant $z_0 := 0$, une expression égale à

$$\prod_{j=1}^{l-1} \lambda \frac{(\lambda(z_j - z_{j-1}))^{a_j - 1}}{(a_j - 1)!} \times \lambda\, e^{-\lambda z_l}, \tag{3.3.3}$$

puisque $a_l = 1$. Il reste donc à intégrer cette densité sur l'évènement A pour trouver la probabilité de A.

Si $Z_1 = X_1$, alors $a_1 = i_1 \geq 1$ et $0 \leq X_1 \leq s_1$; dans le produit (3.3.3), le terme correspondant à $j = 1$ vaut $\lambda \dfrac{(\lambda x_1)^{i_1 - 1}}{(i_1 - 1)!}$. De même, si $Z_j = Y_k$ et $Z_{j+1} = Y_{k'}$ avec $k' > k$, alors $a_j = 1$ et $s_k < Y_k \leq s_{k+1}$; de plus, le terme indicé par j dans (3.3) vaut simplement λ. Enfin, si $Z_j = Y_k$ et $Z_{j+1} = X_{k'}$ avec $k' > k$, alors $k' = k+1$, puis $a_j = 1$, $a_{j+1} = i_{k+1} - 1$ et $s_k < Y_k \leq X_{k+1} \leq s_{k+1}$; de plus, le produit des deux termes consécutifs, indicés par j, $j+1$ vaut : $\lambda^2 \dfrac{(\lambda(x_{k+1} - y_k))^{i_{k+1} - 2}}{(i_{k+1} - 2)!}$. Comme $Z_l = Y_m$ et $s_m < Y_m$, l'intégration de la

densité de Z sur A est égale à :

$$\int_0^{s_1} \lambda \frac{(\lambda x_1)^{i_1-1}}{(i_1-1)!}\,dx_1 \times \prod_{\substack{1\le k\le m-1\\ i_{k+1}\ge 1}} H_k \times \int_{s_m}^{+\infty} \lambda e^{-\lambda y_m}\,dy_m, \qquad (3.3.4)$$

où la première intégrale disparaît si $i_1 = 0$ et où

$$\begin{aligned} H_k &= \iint_{s_k<y_k\le x_{k+1}\le s_{k+1}} \lambda^2 \frac{(\lambda(x_{k+1}-y_k))^{i_{k+1}-2}}{(i_{k+1}-2)!}\,dy_k\,dx_{k+1}, && \text{si } i_{k+1}\ge 2;\\ &= \int_{s_k<y_k\le s_{k+1}} \lambda\,dy_k, && \text{lorsque } i_{k+1}=1. \end{aligned}$$

La première intégrale du produit vaut $(\lambda s_1)^{i_1}/i_1!$, qui donc se réduit à 1 lorsque $i_1 = 0$ et la dernière intégrale est égale à $e^{-\lambda s_m}$. L'intégrale double vaut $H_k = \dfrac{(\lambda t_{k+1})^{i_{k+1}}}{i_{k+1}!}$, d'après le Lemme 3.3.4, une valeur qui est aussi la valeur de H_k lorsque $i_{k+1} = 1$ *et* lorsque $i_{k+1} = 0$.

L'expression (3.3.4) est donc égale à :

$$\frac{(\lambda s_1)^{i_1}}{i_1!} \times \Bigl(\prod_{1\le k\le m-1} \frac{(\lambda t_{k+1})^{i_{k+1}}}{i_{k+1}!}\Bigr) \times e^{-\lambda s_m} = \prod_{1\le k\le m} e^{-\lambda t_k}\frac{(\lambda t_k)^{i_k}}{i_k!},$$

puisque $s_1 = t_1$ et $s_m = t_1 + t_2 + \cdots + t_m$. □

Pour terminer la démonstration du Théorème 3.3.1, il suffit de vérifier que les propriétés (i), (ii) et (iii) de la Définition 3.1.5 sont satisfaites. Or, l'évènement $\{N(0) \ge 1\}$ entraîne que $\{T_1 = 0\}$ s'est produit, mais cet évènement est de probabilité nulle. On a donc $N(0) = 0$, p.s.

Récrivons la Proposition 3.3.3 pour $m = 2$. On obtient :

$$\mathrm{P}\{N(s_1) = i_1, N(s_2) - N(s_1) = i_2\} = \frac{e^{-\lambda t_1}(\lambda t_1)^{i_1}}{i_1!}\frac{e^{-\lambda t_2}(\lambda t_2)^{i_2}}{i_2!}. \qquad (3.3.5)$$

D'après la Proposition 3.3.2, la suite $(\{N(s_1) = i_1\})$ $(i_1 = 0, 1, \dots)$ est un système complet d'évènements. En sommant (3.3.5) par rapport à i_1, de 0 à l'infini, on obtient, pour tout $i_2 \ge 0$,

$$\mathrm{P}\{N(s_2) - N(s_1) = i_2\} = \frac{e^{-\lambda t_2}(\lambda t_2)^{i_2}}{i_2!}. \qquad (3.3.6)$$

D'après la Proposition 3.3.2, on en déduit :

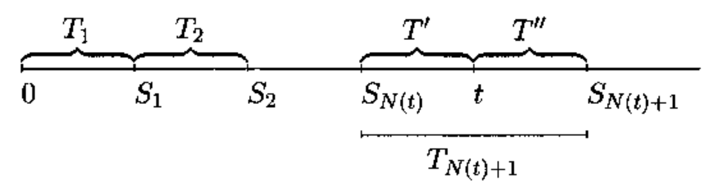
$$\begin{aligned} \mathrm{P}\{N(s_1) = i_1, N(s_2) - N(s_1) = i_2, \dots, N(s_m) - N(s_{m-1}) = i_m\}&\\ = \mathrm{P}\{N(s_1) = i_1\}\,\mathrm{P}\{N(s_2) - N(s_1) = i_2\}&\\ \cdots \times \mathrm{P}\{N(s_m) - N(s_{m-1}) = i_m\}.& \qquad (3.3.7) \end{aligned}$$

Le processus est donc bien à accroissements indépendants. Enfin, la propriété (iii) n'est autre que (3.3.6). □

3.4 ÂGE, TEMPS DE VIE RÉSIDUEL

Considérons un processus de Poisson $\{N(t) : t \geq 0\}$, de densité λ et conservons les notations des précédents paragraphes quant aux variables T_n et S_n.

Pour tout $t > 0$ l'intervalle $[S_{N(t)}, S_{N(t)+1}]$ contient t ; en effet, $S_{N(t)}$ est la date où se produit le *dernier* top avant t (puisque $N(t)$ est une variable aléatoire de Poisson, elle prend, presque sûrement, des valeurs entières positives finies). D'autre part, $S_{N(t)+1}$ est la date où se produit le *premier* top après t.



Partageons l'intervalle $[S_{N(t)}, S_{N(t)+1}]$, de longueur $T_{N(t)+1}$, en les deux intervalles partiels disjoints $[S_{N(t)}, t[$ et $[t, S_{N(t)+1}[$ de longueurs respectives T' et T''. Imaginons que les tops soient les dates de naissance d'individus qui disparaissent au top suivant. Alors l'intervalle $T' = t - S_{N(t)}$ est l'*âge de l'individu en vie à la date t* et $T'' = S_{N(t)+1} - t$ est le *temps de vie résiduelle de l'individu en vie à la date t*. On a $T_{N(t)+1} = T' + T''$; de plus, comme on le montre ci-dessous au § 4.3, les variables aléatoires T' et T'' sont indépendantes.

3.4.1 La loi de T′

Cette variable prend ses valeurs dans $[0, t]$. Pour tout $s > t$, on a donc $\mathrm{P}\{T' > s\} = 0$ et, pour tout $s \in [0, t]$,

$$\begin{aligned}\{T' > s\} &= \{t - S_{N(t)} > s\} = \{S_{N(t)} < t - s\} \\ &= \{0 \text{ top dans } [t-s, t]\};\end{aligned}$$

d'où

$$\begin{aligned}\mathrm{P}\{T' > s\} &= \mathrm{P}\{0 \text{ top dans } [t-s, t]\} \\ &= \mathrm{P}\{0 \text{ top dans } [0, s]\} = e^{-\lambda s}.\end{aligned}$$

Par conséquent,

$$\mathrm{P}\{T' > s\} = \begin{cases} e^{-\lambda s}, & \text{si } s \in [0, t]\,; \\ 0, & \text{si } s > t. \end{cases}$$

La loi de T' est donc la loi de $T_1 \wedge t$.

3.4.2 La loi de T″

Que T'' et T_1 soient de même loi, résulte de la propriété d'absence de mémoire de la loi exponentielle. On peut encore le montrer comme suit : T'' prend ses valeurs dans $[0, +\infty[$. Pour tout $s \geq 0$, on a donc l'équivalence des évènements : $\{T'' > s\} = \{S_{N(t)+1} - t > s\} = \{S_{N(t)+1} > t + s\} = \{0 \text{ top dans } [t, t+s]\}$; d'où $\mathrm{P}\{T'' > s\} = e^{-\lambda s}$. □

3.4.3 Indépendance

Montrons que T' et T'' sont indépendantes. Pour $0 < x < t$, $y > 0$, on a

$$\{T' > x, T'' > y\} = \{0 \text{ top dans } [t-x, t+y]\},$$

d'où

$$\mathrm{P}\{T' > x, T'' > y\} = e^{-\lambda(x+y)} = \mathrm{P}\{T' > x\}\,\mathrm{P}\{T'' > y\}.$$

Pour $x \geq t$, $y > 0$, on a :

$$\mathrm{P}\{T' > x, T'' > y\} = 0 = \mathrm{P}\{T' > x\}\,\mathrm{P}\{T'' > y\}. \quad \square$$

En définitive, on a $T_{N(t)+1} = T' + T''$, où T', T'' sont des variables aléatoires *indépendantes* telles que $\mathcal{L}(T') = \mathcal{L}(T_1 \wedge t)$, $\mathcal{L}(T'') = \mathcal{L}(T_1)$. Il en résulte $\mathcal{L}(T_{N(t)+1}) = \mathcal{L}(T_1 \wedge t) * \mathcal{L}(T_1)$, où "$*$" désigne le produit de convolution (*cf.* [FF1], chap. 11, § 6).

En revanche, les variables aléatoires $T_{N(t)+2}$, $T_{N(t)+3}$, ... sont *indépendantes, identiquement distribuées*, de loi $\mathcal{L}(T_1)$. Il en résulte que la suite des variables aléatoires T'', $T_{N(t)+2}$, $T_{N(t)+3}$, ... qui représente la suite des durées séparant deux tops consécutifs à partir de la date t, a les mêmes propriétés statistiques que la suite des variables aléatoires T_1, T_2, ... qui représente la suite des ces durées à partir de la date 0. Il est alors clair que le nombre de tops dans $]t, t+s]$ $(t, s > 0)$ a la même loi de probabilité que le nombre de tops dans $]0, s]$, autrement dit, que le processus est à accroissements stationnaires.

3.4.4 Temps d'arrêt

Soient $\{N(t) : t \geq 0\}$ un processus de Poisson de densité $\lambda > 0$ et (T_n) $(n \geq 0)$ la suite de ses espacements.

Proposition 3.4.1. *Soit $t > 0$ fixé. Alors*

(a) *la variable aléatoire $N(t)$ n'est pas un temps d'arrêt adapté à la suite (T_n) $(n \geq 1)$;*

(b) *la variable $N(t)+1$ est un temps d'arrêt adapté à la suite (T_n) $(n \geq 1)$.*

Démonstration. L'évènement $\{N(t) = n\}$ est la conjonction des évènements $\{T_1 + \cdots + T_n \leq t\}$ et $\{T_1 + \cdots + T_{n+1} > t\}$. Ainsi, $N(t)$ est l'indice du *dernier* top se produisant dans $[0, t]$. En revanche, $\{N(t) + 1 = n\} = \{N(t) = n - 1\}$, qui est la conjonction des évènements $\{T_1 + \cdots + T_{n-1} \leq t\}$ et $\{T_1 + \cdots + T_n > t\}$, est l'indice du premier top se produisant dans $]t, +\infty[$. $\square$

Revenons à l'étude des instants $S_{N(t)}$, $S_{N(t)+1}$ et des variables aléatoires $T' = t - S_{N(t)}$, $T'' = S_{N(t)+1} - t$. On a : $\mathbb{E}[T'] = \int_0^t \mathrm{P}\{T' > s\}\,ds = \int_0^t e^{-\lambda s}\,ds = (1 - e^{-\lambda t})/\lambda$. D'où $\mathbb{E}[S_{N(t)}] = t - (1 - e^{-\lambda t})/\lambda$; puis $\mathbb{E}[T''] = 1/\lambda$, d'où $\mathbb{E}[S_{N(t)+1}] = t + 1/\lambda$.

Si $N(t)$ était un temps d'arrêt, on pourrait appliquer l'identité de Wald généralisée (*cf.* § 2.5) et l'on aurait : $\mathbb{E}[S_{N(t)}] = \mathbb{E}[N(t)]\,\mathbb{E}[T_1] = \lambda t \cdot (1/\lambda) = t$. On ne trouve pas le même résultat !

En revanche, $N(t)+1$ est un temps d'arrêt. On peut appliquer l'identité de Wald. On obtient : $\mathbb{E}[S_{N(t)+1}] = \mathbb{E}[N(t)+1]\,\mathbb{E}[T_1] = (\lambda t+1)/\lambda = t + 1/\lambda$.

3.4.5 Le paradoxe de l'inspection

On peut l'exprimer de la façon suivante. On dispose d'une provision infinie d'ampoules dont les temps de vie sont des variables aléatoires *indépendantes*, identiquement distribuées selon la loi exponentielle de paramètre $\lambda > 0$, avec λ inconnu. À l'instant 0, on allume une ampoule et on la remplace instantanément dès qu'elle s'éteint. La suite des instants où les ampoules s'éteignent (les tops) définit alors un processus de Poisson $\{N(t) : t \geq 0\}$, où $N(t)$ désigne le nombre d'ampoules consommées dans l'intervalle $[0,t]$.

On note T_1, T_2, ... les temps de vie des ampoules successives. Ce sont des variables aléatoires indépendantes suivant toutes la loi exponentielle de paramètre inconnu λ. On se propose d'estimer λ au moyen de la procédure suivante (« inspection ») :

1) On fixe un instant $t > 0$.

2) On observe le temps de vie de l'ampoule qui brûle à l'instant t. Avec les notations ci-dessus, ceci revient à étudier la variable aléatoire $T_{N(t)+1}$. Or $T_{N(t)+1} = T' + T''$, où T' et T'' sont deux variables aléatoires *indépendantes* dont les lois de probabilité ont été calculées ci-dessus). On a donc

$$\mathbb{E}[T'] = \int_0^t \mathrm{P}\{T' > s\}\,ds = \int_0^t e^{-\lambda s}\,ds = \frac{1-e^{-\lambda t}}{\lambda}\,;$$

$$\mathbb{E}[T''] = \frac{1}{\lambda}.$$

Il en résulte :

$$\mathbb{E}[T_{N(t)+1}] = \mathbb{E}[T'] + \mathbb{E}[T''] = \frac{2}{\lambda} - \frac{e^{-\lambda t}}{\lambda} > \frac{1}{\lambda} = \mathbb{E}[T_1].$$

On constate que pour tout $t > 0$ la durée de vie moyenne de l'ampoule *qu'on inspecte à l'instant t* est strictement supérieure à la durée de vie moyenne des ampoules non inspectées. C'est en ce phénomène que réside le *paradoxe de l'inspection.* L'inspection, i.e., l'observation de l'ampoule à l'instant t produit une perturbation qui a pour effet de *dilater* la durée de vie moyenne de l'ampoule inspectée (*cf.* le proverbe italien : « L'œil du maître engraisse le cheval. »).

Remarque 1. Lorsque t tend vers l'infini, la loi de probabilité de T' tend vers celle de T_1 et $\mathbb{E}[T_{N(t)+1}] \to 2/\lambda = 2\,\mathbb{E}[T_1]$. Pour de grandes valeurs de t la durée de vie moyenne de l'ampoule qu'on observe à l'instant t est pratiquement *le double* de la durée de vie moyenne des autres ampoules. Le paradoxe de l'inspection est particulièrement saisissant dans ce cas.

Remarque 2. Comme $T_{N(t)+1} = T' + T''$, la loi de probabilité de $T_{N(t)+1}$ est la convolution de la loi exponentielle de paramètre λ et de la loi de probabilité $T_1 \wedge t$. Lorsque t tend vers l'infini, cette loi tend vers la convolution de deux lois exponentielles de paramètre λ, c'est-à-dire, vers la loi $\Gamma(2,\lambda)$.

Remarque 3. Parmi les ampoules, il y en a qui brûlent plus longtemps que les autres. Si l'on choisit un instant $t > 0$, on a plus de chances de tomber sur une ampoule qui brûle longtemps que sur une ampoule qui brûle moins longtemps ; ce serait là une explication intuitive du paradoxe. Il y a pourtant un hic dans cet argument, puisqu'il n'exclut pas qu'il puisse y avoir des ampoules qui brûlent peu de temps ; l'explication n'est valable qu'*a posteriori*.

3.5 DISTRIBUTION CONDITIONNELLE DES INSTANTS D'ARRIVÉE

Supposons que nous connaissions le nombre de tops qui se sont produits dans $[0, t]$. Que peut-on dire de la distribution conditionnelle de leurs instants d'arrivée ?

Théorème 3.5.1. *Conditionnellement à l'évènement $\{N(t) = n\}$, la loi de probabilité des n tops apparaissant dans $[0, t]$ est égale à la loi de probabilité de n variables aléatoires indépendantes, identiquement distribuées, de loi uniforme sur $[0, t]$.*

On peut écrire l'énoncé du précédent théorème sous la forme suivante.

Théorème 3.5.1′. *Soient I un intervalle de $[0, t]$ et n un entier. Soit, d'autre part, $(Y_1, \ldots, Y_n)$ une suite de n variables aléatoires, indépendantes, identiquement distribuées de loi uniforme sur $[0, t]$. Alors, conditionnellement à l'évènement $\{N(t) = n\}$, la probabilité pour qu'exactement k tops se produisent dans I est égale à la probabilité pour qu'exactement k parmi les variables $Y_1, \ldots, Y_n$ prennent leurs valeurs dans I.*

Démonstration. Supposons que I soit de longeur u $(0 \le u \le t)$ et désignons par X (resp. Y) le nombre d'instants-τ situés dans I (resp. dans $[0, t] \setminus I$). On a $X + Y = N(t)$; d'autre part, X et Y sont des variables aléatoires indépendantes suivant des lois de Poisson de paramètre respectif λu et $\lambda(t-u)$. Pour $k = 0, 1, \ldots, n$, on a alors :

$$\begin{aligned}
\mathrm{P}\{X = k \mid N(t) = n\} &= \mathrm{P}\{X = k \mid X + Y = n\} \\
&= \frac{\mathrm{P}\{X = k, X + Y = n\}}{\mathrm{P}\{N(t) = n\}} = \frac{\mathrm{P}\{X = k, Y = n - k\}}{\mathrm{P}\{N(t) = n\}} \\
&= \frac{\mathrm{P}\{X = k\}\mathrm{P}\{Y = n - k\}}{\mathrm{P}\{N(t) = n\}} \\
&= e^{-\lambda u}\frac{(\lambda u)^k}{k!} e^{-\lambda(t-u)} \frac{(\lambda(t-u))^{n-k}}{(n-k)!} \Big/ e^{-\lambda t}\frac{(\lambda t)^n}{n!} \\
&= \binom{n}{k}\Big(\frac{u}{t}\Big)^k\Big(1 - \frac{u}{t}\Big)^{n-k},
\end{aligned}$$

une expression qui est indépendante de λ, qui est bien la probabilité pour qu'exactement k variables parmi $Y_1, \ldots, Y_n$ prennent leurs valeurs dans I. □

Donnons maintenant un lemme sur les échantillons ordonnés. Si $X = (X_1, \dots, X_n)$ est une suite de n variables aléatoires indépendantes, identiquement distribuées, absolument continues (de sorte que la probabilité pour que deux variables X_i et X_j $(i \neq j)$ soient égales, est nulle), on note $(X_1^* < X_2^* < \cdots < X_n^*)$ la suite des X_i rangés en ordre croissant. Cette suite est appelée *l'échantillon ordonné.*

Lemme 3.5.2. *Soit $X = (X_1, \dots, X_n)$ une suite de n variables aléatoires, indépendantes, identiquement distribuées, absolument continues, de densité f. Alors l'échantillon ordonné $X^* := (X_1^* < X_2^* < \cdots < X_n^*)$ admet une densité conjointe donnée par*

$$n! \prod_{i=1}^{n} f(x_i)\, I_E \quad avec \quad E = \{(x_1, \dots, x_n) : x_1 < x_2 < \cdots < x_n\}.$$

Démonstration. À tout élément σ appartenant au groupe des $n!$ permutations de $1, 2, \dots, n$, faisons correspondre le sous-ensemble $E(\sigma)$ des vecteurs $(x_1, x_2, \dots, x_n)$ de $\mathbb{R}^n$ tels que $x_{\sigma 1} < x_{\sigma 2} < \cdots < x_{\sigma n}$. Notons que $\mathbb{R}^n$ est la réunion des ensembles *disjoints* $E(\sigma)$ et du sous-ensemble, de mesure de Lebesgue nulle, des vecteurs qui ont au moins deux composantes égales.

Soit g la fonction qui envoie tout vecteur sur le réarrangement croissant de ses composantes. En particulier, $X^* = g \circ X$. De plus, la restriction g_σ de g à $E(\sigma)$ envoie le vecteur $(x_1, x_2, \dots, x_n)$ de $E(\sigma)$ sur le vecteur $(x_{\sigma 1}, x_{\sigma 2}, \dots, x_{\sigma n})$ de E ; en fait, g_σ est une transformation linéaire ayant pour matrice une matrice de permutation, dont le déterminant est ± 1. Le jacobien de g_σ est donc égal à 1. Alors, pour toute permutation σ et tout borélien $B \subset E$, on a, par le changement de variables $x \mapsto g_\sigma(x)$, en posant $f_X(x) := f(x_1) \cdots f(x_n)$,

$$\int_{g_\sigma^{-1}(B)} f_X(x)\, dx = \int_B f_X(g_\sigma^{-1}(y))\, dy.$$

Or, si τ est la permutation inverse de σ, pour $y = (y_1 < \cdots < y_n)$, on a : $f_X(g_\sigma^{-1}(y)) = f(y_{\tau 1}) \cdots f(y_{\tau n}) = f(y_1) \cdots f(y_n) = f_X(y)$ et donc

$$\int_{g_\sigma^{-1}(B)} f_X(x)\, dx = \int_B f_X(y)\, dy.$$

Comme $\mathrm{P}_{X^*}(B) = \mathrm{P}\{X^* \in B\} = \mathrm{P}\{g \circ X \in B\} = \mathrm{P}_X\{g \in B\}$, on en déduit :

$$\begin{aligned} \mathrm{P}_{X^*}(B) &= \int_{\mathbb{R}^n} I_{\{g \in B\}} f_X(x)\, dx = \sum_\sigma \int_{E(\sigma)} I_{\{g \in B\}} f_X(x)\, dx. \\ &= \sum_\sigma \int_{g_\sigma^{-1}(B)} f_X(x)\, dx = \int_B n!\, f_X(y)\, dy, \end{aligned}$$

ce qui montre que la densité de X^* est $n!\, f(x_1) \cdots f(x_n)$ sur E et 0 ailleurs. □

Cas particulier. Si le n-uple $(X_1, X_2, \dots, X_n)$ est formé de variables aléatoires *indépendantes*, identiquement distribuées, de loi uniforme sur $[0, t]$, la densité

conjointe de l'échantillon ordonné $(X_1^*, X_2^*, \cdots, X_n^*)$ est donnée par

$$n!\,\frac{1}{t^n}\,I_E \quad \text{avec} \quad E = \{(x_1, \ldots, x_n) : 0 < x_1 < x_2 < \cdots < x_n < t\}.$$

Proposition 3.5.3. *Conditionnellement à l'évènement $\{N(t) = n\}$, le n-uple $(S_1, S_2, \ldots, S_n)$ a même loi de probabilité que le n-uple ordonné correspondant à n variables aléatoires indépendantes, identiquement distribuées, de loi uniforme sur $[0, t]$.*

Démonstration. Soient $0 < x_1 < x_2 < \cdots < x_n < t$ une suite (strictement) croissante et $h_1, h_2, \ldots, h_n$ des nombres strictement positifs suffisamment petits pour que $x_1 \leq x_1 + h_1 < x_2 \leq x_2 + h_2 < \cdots < x_{n-1} \leq x_{n-1} + h_{n-1} < x_n \leq x_n + h_n < t$. Alors

$$\begin{aligned}
\mathrm{P}\{S_1 \in [x_1, x_1+h_1], \ldots, S_n \in [x_n, x_n+h_n] \,|\, N(t) = n\} \\
&= \frac{\mathrm{P}\{S_1 \in [x_1, x_1+h_1], \ldots, S_n \in [x_n, x_n+h_n], N(t) = n\}}{\mathrm{P}\{N(t) = n\}} \\
&= \frac{\mathrm{P}\{\text{un seul top dans } [x_i, x_i+h_i]\ (i = 1, \ldots, n);\ 0 \text{ ailleurs}\}}{\mathrm{P}\{N(t) = n\}} \\
&= \frac{e^{-\lambda h_1}\lambda h_1 \cdots e^{-\lambda h_n}\lambda h_n\, e^{-\lambda(t-h_1-\cdots-h_n)}}{e^{-\lambda t}\dfrac{(\lambda t)^n}{n!}} = n!\,\frac{1}{t^n} h_1 \cdots h_n.
\end{aligned}$$

En divisant par $h_1, \ldots, h_n$ et en faisant tendre successivement $h_1, \ldots, h_n$ vers 0, on trouve la densité correspondante :

$$n!\,\frac{1}{t^n} I_E, \quad \text{avec} \quad E = \{(x_1, \ldots, x_n) : 0 < x_1 < x_2 < \cdots < x_n < t\}. \quad \square$$

3.6 CARACTÉRISATION « QUALITATIVE » D'UN PROCESSUS DE POISSON

Comme déjà annoncé dans le premier paragraphe, quatre conditions "qualitatives" suffisent à s'assurer qu'un processus de comptage est de Poisson.

Théorème 3.6.1. *Soit $\{N(t) : t \geq 0\}$ un processus de comptage vérifiant les propriétés suivantes :*

(1) $N(0) = 0$;

(2) *le processus est à accroissements stationnaires ;*

(3) *le processus est à accroissements indépendants ;*

(4) *le processus est localement continu en probabilité, c'est-à-dire, pour tout $t \geq 0$, $\mathrm{P}\{N(t+h) - N(t) \geq 1\}$ tend vers 0 avec h.*

Alors le processus est de Poisson.

Démonstration. En vertu de (2) et de (1), pour tout $h > 0$, on a : $\mathrm{P}\{N(t+h) - N(t) \geq 1\} = \mathrm{P}\{N(h) - N(0) \geq 1\} = \mathrm{P}\{N(h) \geq 1\}$, une quantité qu'on pose égale à $p(h)$. C'est la probabilité pour qu'il se produise au moins un top dans un intervalle de temps de longueur $h > 0$. La condition (4) implique $\lim_{h \to 0} p(h) = 0$.

Pour terminer la démonstration du théorème, trois lemmes sont à établir.

Lemme 3.6.2. *Pour tout $t \geq 0$, on a : $p(t) = 1 - e^{-\lambda t}$, avec $\lambda \geq 0$.*

Démonstration. Posons $f(t) := 1 - p(t) = \mathrm{P}\{N(t) = 0\}$. Il résulte des conditions (2) et (3) que, pour tout $t > 0$ et tout $h > 0$, on a : $f(t+h) = f(t)f(h)$. La solution identiquement nulle est exclue; en effet, cette solution équivaudrait à $p(t) = 1$ pour tout $t \geq 0$, en contradiction avec (4). On a donc $f(t) = e^{-\lambda t}$, avec, puisque f est bornée, $\lambda \geq 0$ (*cf.* [FF1], chap. 14, Exercice 1), d'où $p(t) = 1 - f(t) = 1 - e^{-\lambda t}$, avec $\lambda \geq 0$. Le cas $\lambda = 0$ est inintéressant, il entraînerait que, pour tout $t \geq 0$, on aurait $p(t) = 0$, c'est-à-dire, qu'avec une probabilité égale à 1, il ne se produirait aucun top. □

Dans la suite, on suppose donc $\lambda > 0$ et, par conséquent,

$$p(t) \sim \lambda t \quad (t \downarrow 0). \tag{3.6.1}$$

Lemme 3.6.3. *Pour tout intervalle de temps $]a, b]$, $-\infty < a < b < +\infty$, la probabilité pour qu'il se produise une infinité de tops dans $]a, b]$ est nulle.*

Démonstration. Si, avec une probabilité égale à 1, il y avait une infinité de tops dans $]a, b]$, alors, avec une probabilité égale à 1, il y aurait une infinité de tops dans l'un au moins des intervalles $]a, c]$, $]c, b]$, avec $c = (a+b)/2$, disons dans $]a, c]$. En vertu de (2), avec une probabilité égale à 1, il y aurait une infinité de tops dans $]c, b]$. Par dichotomies successives, on voit qu'avec une probabilité égale à 1, chaque point de $]a, b]$ serait point d'accumulation d'instants-τ, donc pour tout $h > 0$, on aurait $P(h) = 1$, en contradiction avec (4). □

Lemme 3.6.4. *Le nombre $N(t)$ de tops se produisant dans $[0, t]$ suit la loi de Poisson de paramètre λt.*

Démonstration. Partageons l'intervalle $[0, t]$ en n intervalles partiels égaux, au moyen des points de subdvision : $t_1 := 0$, $t_1 := t/n$, $t_2 := 2t/n$, ... , $t_{n-1} := (n-1)t/n$, $t_n := t$. Soient $I_k :=]t_{k-1}, t_k]$ le $k^{\text{ième}}$ intervalle partiel et X_k le nombre de tops se produisant dans I_k $(1 \leq k \leq n)$. D'après (3), les variables aléatoires X_k sont *indépendantes*, prenant des valeurs entières positives. De plus, d'après (2), elles sont *identiquement distribuées.*

Associons à l'intervalle I_k la variable aléatoire Y_k définie par $Y_k := I_{\{X_k \geq 1\}}$. Les Y_k sont des variables aléatoires de Bernoulli, indépendantes, identiquement distribuées, de loi commune $(1 - p(t/n))\varepsilon_0 + p(t/n)\varepsilon_1$. Posons $S_n := Y_1 + \cdots + Y_n$. On a naturellement $S_n \leq N(t)$. En général, $S_n < N(t)$, mais, pour n suffisamment grand, on peut s'attendre à ce que $S_n = N(t)$. En fait, nous montrons que :

$$S_n \xrightarrow{p.s.} N(t) \quad (n \to +\infty). \tag{3.6.2}$$

D'après le Lemme 3.6.3, il y a une probabilité nulle pour qu'il se produise une infinité de tops dans $[0,t]$; il y a également une probabilité nulle pour que deux tops se produisent au même instant. Par conséquent, pour presque toute réalisation ω du processus, il y a, dans $[0,t]$, seulement un nombre *fini* d'instants-τ, tous *distincts*. On peut donc, pour cette réalisation ω, choisir un entier n_0 assez grand pour que, pour tout $n \geq n_0$, chacun des intervalles partiels $I_1, \ldots, I_k$ contienne *au plus* un seul instant-τ. Ainsi, pour cette réalisation ω et pour tout $n \geq n_0$, on a : $S_n(\omega) = N(t)(\omega)$, ce qui entraîne (3.6.2).

Comme la convergence presque sûre entraîne la convergence en loi, on a aussi $S_n \xrightarrow{\mathcal{L}} N(t)$ $(n \to +\infty)$. Cette propriété permet de déterminer la loi de $N(t)$. En effet, S_n est une variable aléatoire binomiale de paramètre $(n, p(t/n))$. On a, d'autre part, $p(t/n) \to 0$ et $\mathbb{E}[S_n] = n\,p(t/n)$, une quantité qui est équivalente à $n\,\lambda t/n = \lambda\,t$, d'après (3.6.1). D'où, $\mathbb{E}[S_n] \to \lambda\,t$. Il en résulte que S_n converge en loi vers une variable aléatoire de Poisson de paramètre $\lambda\,t$. Ainsi, sa limite $N(t)$ suit cette loi de Poisson. □

L'hypothèse (2) de stationnarité permet alors de conclure que le nombre de tops se produisant dans *tout* intervalle de longueur t suit une loi de Poisson de paramètre $\lambda\,t$. Ceci achève la démonstration du Théorème 3.6.1.

3.7 COMPLÉMENTS ET EXERCICES

3.1 Soient $\{N(t) : t \geq 0\}$ un processus de Poisson de densité $\lambda > 0$ et T_1 l'instant où se produit le premier top. Montrer directement que la loi de probabilité de T_1, conditionnellement à l'évènement $\{N(t) = 1\}$, est la loi uniforme sur $[0,t]$.

3.2 Des greffons de foie arrivent à un bloc opératoire suivant un processus de Poisson de densité λ. Deux patients attendent d'être greffés, le premier (resp. le second) a une durée de vie qui suit une loi exponentielle de paramètre μ_1 (resp. μ_2). Il est entendu que le premier greffon qui arrive au bloc est pour le premier patient, s'il est encore en vie, sinon au second, s'il est encore en vie. La probabilité que le premier (resp. le second) patient soit greffé est : $\dfrac{\lambda}{\lambda+\mu_1}$ (resp. $\dfrac{\lambda\mu+\lambda^2}{(\lambda+\mu_2)(\lambda+\mu_1+\mu_2)}$).

3.3 Les employés d'une entreprise arrivent au travail selon un processus de Poisson $\{N(t) : t \geq 0\}$ de densité $\lambda > 0$. Pour $i = 1, 2, \ldots$, soit S_i la date d'arrivée du $i^{\text{ième}}$ employé.

(a) Le nombre total des heures de travail effectuées dans l'entreprise jusqu'à la date $t \geq 0$ est : $X(t) = \sum_{i=1}^{N(t)} (t - S_i)$.

(b) Utiliser la Proposition 3.5.3 pour évaluer $\mathbb{E}[\sum_{i=1}^{N(t)} S_i \mid N(t) = n]$. En déduire : $\mathbb{E}[X(t)] = \lambda t^2/2$.

3.4 Un compteur reçoit des impulsions selon un processus de Poisson $\{N(t) : t \geq 0\}$, de densité $\lambda > 0$. On suppose que l'amplitude d'une impulsion décroît avec le temps, suivant un taux exponentiel, c'est-à-dire que si une impulsion a une amplitude A à la date 0, alors son amplitude à la date $t > 0$ est $A\,e^{-\alpha t}$, avec $\alpha > 0$. On suppose, en outre, que les amplitudes initiales des différentes impulsions sont indépendantes et suivent toutes la même loi qu'une variable aléatoire A, ayant une espérance mathématique $\mathbb{E}[A]$ finie. Elles sont, enfin, indépendantes du processus $\{N(t) : t \geq 0\}$. On désigne par $X_1, X_2, \ldots$ la suite des instants d'arrivée des impulsions et par $A_1, A_2, \ldots$ la suite des amplitudes initiales respectives. Alors l'amplitude totale mesurée à la date $t > 0$ est donnée par

$$A(t) = \sum_{k=1}^{N(t)} A_k\, e^{-\alpha(t - X_k)}.$$

Calculer $\mathbb{E}[A(t) \mid N(t) = n]$, puis $\mathbb{E}[A(t)]$.

3.5 Soit $\{N(t) : t \geq 0\}$ un processus de Poisson, de densité $\lambda > 0$. Calculer $\mathrm{Cov}(N(s), N(t))$ pour $s, t \geq 0$.

3.6 Un processus de Poisson est localement continu en moyenne quadratique.

3.7 En tout point t, un processus de Poisson est dérivable en probabilité avec dérivée nulle. En revanche, il n'est pas dérivable en moyenne quadratique.

3.8 Soit $\{N(t) : t \geq 0\}$ un processus de Poisson de densité $\lambda > 0$ et T une variable aléatoire, indépendante de ce processus, de loi exponentielle, de paramètre $\mu > 0$. On désigne par Z le nombre de tops du processus qui se produisent dans l'intervalle aléatoire $[0, T]$. Déterminer la loi de probabilité de Z. (Dans la terminologie de l'Exercice 3.2, la variable Z est le nombre de greffons qui arrivent tant que le premier patient est vivant. La probabilité que le premier patient soit greffé est $\mathrm{P}\{Z \geq 1\}$.)

3.9 Soient $\{N(t) : t \geq 0\}$ un processus de Poisson de densité $\lambda > 0$ et X_0 une variable aléatoire, indépendante de ce processus, de loi $\frac{1}{2}(\varepsilon_{-1} + \varepsilon_{+1})$. On pose $X(t) := X_0\,(-1)^{N(t)}$. Calculer :

(a) $\mathbb{E}[X(t)]$ $(t \geq 0)$; (b) $\mathrm{Cov}(X(s), X(t))$ $(0 < s < t < +\infty)$.

3.10 Soit $(X_1^*, \ldots, X_n^*)$ l'échantillon ordonné d'une suite $(X_1, \ldots, X_n)$ de variables aléatoires, indépendantes, uniformément distribuées sur $[0, t]$. D'après la Proposition 3.5.3, cet échantillon ordonné a pour densité conjointe $n!/t^n$ sur l'ensemble $\{0 < x_1 < \cdots < x_n < t\}$ et 0 ailleurs.

(a) Pour $1 \le s \le n$, la densité $f_{X_s^*}(x_s)$ de X_s^* est donnée par

$$\binom{n}{s-1, 1, n-s} \frac{1}{t^n} x_s^{s-1} (t - x_s)^{n-s} \quad (0 < x_s < t);$$

et la densité $f_{Y_s}(y_s)$ de $Y_s := t - X_s^*$ par

$$\binom{n}{n-s, 1, s-1} \frac{1}{t^n} (t - y_s)^{s-1} y_s^{n-s} \quad (0 < y_s < t).$$

(b) Pour $1 \le r < s \le n$, la densité $f_{X_r^*, X_s^*}(x_r, x_s)$ du couple (X_r^*, X_s^*) est donnée par :

$$\binom{n}{r-1, 1, s-r-1, 1, n-s} \frac{1}{t^n} x_r^{r-1} (x_s - x_r)^{s-r-1} (t - x_s)^{n-s} \qquad (0 < x_r < x_s < t).$$

(c) Avec la convention $X_0^* := 0$, $X_{n+1}^* := t$, pour $0 \le r < s \le n+1$ et $(r, s) \ne (0, n+1)$, on pose $U_{r,s} := X_s^* - X_r^*$. Alors la densité $f_{U_{r,s}}(u)$ de $U_{r,s}$ est donnée par :

$$\binom{n}{s-r-1, 1, n-s+r} \frac{1}{t^n} u^{s-r-1} (t - u)^{n-s+r} \quad (0 < u < t).$$

(d) La densité de l'*étendue* $X_n^* - X_1^*$ de l'échantillon $(X_1, \ldots, X_n)$ est donnée par :

$$n(n-1) \frac{1}{t^n} u^{n-2} (t - u) \quad (0 < u < t).$$

(e) Toutes les variables aléatoires $U_{r,r+1}$ $(0 \le r \le n)$ $(n \ge 2)$ ont *même* densité donnée par

$$n \, \frac{1}{t^n} \, (t - u)^{n-1} \quad (0 < u < t);$$

donc même fonction de survie :

$$\mathrm{P}\{U_{r,r+1} > u\} = \frac{1}{t^n} (t - u)^n \quad (0 < u < t).$$

3.11 Soit $\{N(t) : t \ge 0\}$ un processus de Poisson de densité $\lambda > 0$. Les T_n, S_n conservent la même signification que dans tout le chapitre. On suppose qu'un top ne peut être enregistré, que si la longueur de l'intervalle de temps, qui le sépare du top qui le suit immédiatement, est au moins égale à une certaine quantité $l \ge 0$ (inertie de l'appareil enregistreur). On désigne par $Q(t)$ le nombre de tops survenus dans $[0, t]$ $(t > 0)$ et *qui ont été enregistrés.*

(a) Pour tout $t > 0$, le nombre $Q(t)$ a même loi que $\sum_{k=1}^{N(t)} I_{\{T_k > l\}}$.

(b) Conditionnellement à l'évènement $\{N(t) = n\}$, les variables $T_1, \ldots, T_n$ ont même loi que les variables notées $U_{0,1}, \ldots, U_{n-1,n}$ dans l'Exercice 3.10 (e). En déduire l'expression de $\mathbb{E}[\sum_{k=1}^{N(t)} I_{\{T_k > l\}} \,|\, N(t) = n]$.

(c) Calculer $\mathbb{E}[Q(t)]$. Étudier la cas particulier $l = 0$, $l = t$.

Chapitre 4

Applications des processus de Poisson

Dans ce chapitre, nous nous proposons de donner d'autres propriétés des processus de Poisson, ainsi que plusieurs applications. Les notations du chapitre précédent sont conservées. En particulier, le processus de Poisson est noté $\{N(t) : t \geq 0\}$; de plus, les variables T_n et S_n conservent les mêmes significations.

4.1 PROCESSUS DE POISSON MARQUÉS

On suppose donnés :

(1) un processus de Poisson $\{N(t) : t \geq 0\}$, de densité $\lambda > 0$;

(2) un ensemble fini $\{1, 2, \ldots, r\}$ ($r \geq 2$), l'ensemble des *types* ;

(3) pour tout $s \geq 0$, une loi de probabilité $g(s)$ portée par l'ensemble $\{1, 2, \ldots, r\}$, c'est-à-dire la donnée d'une suite $(g_1(s), \ldots, g_r(s))$ de nombres réels, positifs, tels que $g_1(s) + \cdots + g_r(s) = 1$.

Si un top du processus de Poisson se produit à l'instant $s > 0$, alors indépendamment de tous les autres tops, on le classe en type 1, 2, ... , r, avec la probabilité $g_1(s)$, $g_2(s)$, ... , $g_r(s)$, respectivement. On note $N_1(t)$, $N_2(t)$, ... , $N_r(t)$, le nombre de tops de type 1, 2, ..., r se produisant dans $[0, t]$, respectivement. On a naturellement $N(t) = N_1(t) + N_2(t) + \cdots + N_r(t)$. Les processus $\{N_i(t) : t \geq 0\}$ ($i = 1, 2, \ldots, r$) sont dits processus *marqués* (par le classement ainsi proposé). Il s'agit de préciser justement ce classement et de voir dans quelles conditions les processus marqués ainsi obtenus sont encore de Poisson.

Proposition 4.1.1. *Avec les hypothèses précédentes, pour $k_1 \geq 0, \ldots, k_r \geq 0$ tels que $k_1 + \cdots + k_r = n$, on a :*

$$\begin{aligned}&\mathrm{P}\{N_1(t) = k_1, \ldots, N_r(t) = k_r\}\\ &\qquad = \mathrm{P}\{N_1(t) = k_1, \ldots, N_r(t) = k_r \mid N(t) = n\}\, \frac{e^{-\lambda t}(\lambda t)^n}{n!}. \quad (4.1.1)\end{aligned}$$

Démonstration. Pour le système complet d'évènements $\{N(t) = n\}$ $(n \geq 0)$, la somme $\sum_{n \geq 0} \mathrm{P}\{N_1(t) = k_1, \ldots, N_r(t) = k_r, N(t) = n\}$ se réduit au seul terme pour lequel $k_1 + \cdots + k_r = n$, qui est encore égal au membre de droite de l'identité (4.1.1). □

La proposition précédente montre que pour aller plus loin dans le calcul de la distribution du vecteur $(N_1(t), \ldots, N_r(t))$, il faut connaître la loi de probabilité conditionnelle $\mathrm{P}\{\cdot \mid N(t) = n\}$. Celle-ci dépend évidemment de la *procédure* de classement des tops.

Prenons, tout d'abord, $r = 2$ (types) et pour loi $g(s)$ la loi (constante), $p\varepsilon_1 + q\varepsilon_2$, qui *ne* dépend *pas* de s. Dans ce cas, la loi de probabilité conditionnelle $\mathrm{P}\{\cdot \mid N(t) = n\}$ est simplement la loi binomiale $B(n, p)$. On en tire, puisque $n = k + l$,

$$\begin{aligned}\mathrm{P}\{N_1(t) = k, N_2(t) = l\} &= \binom{k+l}{k} p^k q^l \cdot e^{-\lambda t} \frac{(\lambda t)^{k+l}}{(k+l)!}\\ &= e^{-\lambda tp} \frac{(\lambda tp)^k}{k!} e^{-\lambda tq} \frac{(\lambda tq)^l}{l!}.\end{aligned}$$

On en déduit les lois marginales, qui valent

$$\mathrm{P}\{N_1(t) = k\} = \sum_{l \geq 0} \mathrm{P}\{N_1(t) = k, N_2(t) = l\} = e^{-\lambda tp} \frac{(\lambda tp)^k}{k!},$$

$$\mathrm{P}\{N_2(t) = l\} = \sum_{k \geq 0} \mathrm{P}\{N_1(t) = k, N_2(t) = l\} = e^{-\lambda tq} \frac{(\lambda tq)^l}{l!},$$

ce qui montre que $N_1(t)$ et $N_2(t)$ suivent des lois de Poisson de paramètres λtp, λtq.

Finalement, $\mathrm{P}\{N_1(t) = k, N_2(t) = l\} = \mathrm{P}\{N_1(t) = k\}\mathrm{P}\{N_2(t) = l\}$, ce qui montre que les processus $\{N_1(t) : t \geq 0\}$ et $\{N_2(t) : t \geq 0\}$ sont indépendants. Le théorème suivant a donc été établi.

Théorème 4.1.2. *Si chaque top d'un processus de Poisson de densité λ est classé en type* 1 *ou* 2, *avec les probabilités p et q, indépendamment des autres tops, alors les processus marqués $\{N_1(t) : t \geq 0\}$ des tops de type* 1 *et $\{N_2(t) : t \geq 0\}$ des tops de type* 2 *sont eux-mêmes des processus de Poisson, indépendants, de densités respectives λp et λq.*

Le Théorème 4.1.2 admet la généralisation suivante dont nous laissons la démonstration au lecteur.

Théorème 4.1.3. *Soient $\{N(t) : t \geq 0\}$ un processus de Poisson de densité $\lambda > 0$ et r un entier tel que $r \geq 2$. Supposons qu'à chaque instant où un top arrive, on le classe en top de type 1, ... , de type r, avec les probabilités respectives $p_1, \ldots, p_r$ $(p_1 + \cdots + p_r = 1)$, indépendamment des autres tops. Alors les processus marqués $\{N_1(t) : t \geq 0\}, \ldots, \{N_r(t) : t \geq 0\}$ sont encore des processus de Poisson, de densités respectives $\lambda p_1, \ldots, \lambda p_r$. Ils sont, en outre, indépendants.*

Exemples

(1) Le nombre d'objets produits par une machine-outil peut être considéré comme un processus de Poisson $\{N(t) : t \geq 0\}$, de densité $\lambda > 0$. Ces objets sont produits indépendamment les uns des autres et la probabilité pour qu'un objet soit défectueux est p $(0 < p < 1)$. Alors le nombre d'objets défectueux produits par la machine-outil est un processus de Poisson de densité $\lambda p > 0$. De façon analogue, le nombre d'objets conformes produits par la machine est un processus de Poisson de densité $\lambda q > 0$ et ces deux processus sont indépendants.

(2) Soit $N(t)$ le nombre de sinistres enregistrés par une compagnie d'assurance durant l'intervalle de temps $[0, t]$. Le processus $\{N(t) : t \geq 0\}$ peut être considéré comme un processus de Poisson de densité $\lambda > 0$. Supposons que chaque sinistre est classé en l'une des deux catégories (types) :

type 1 : le coût du sinistre est supérieur à 5.000 euros;

type 2 : le coût du sinistre est au plus égal à 5.000 euros.

Supposons, en outre, que la proportion des sinistres de type 1 est p $(0 < p < 1)$, celle des sinistres de type 2, $q = 1 - p$ et que les sinistres soient indépendants les uns des autres. Désignons par $N_1(t)$, $N_2(t)$ le nombre des sinistres de type 1 et 2 dans l'intervalle de temps $[0, t]$, respectivement. Alors les processus $\{N_1(t) : t \geq 0\}$ et $\{N_2(t) : t \geq 0\}$ sont des processus de Poisson indépendants, de densités respectives λp, λq.

(3) Dans l'exemple précédent, au lieu de ne considérer que deux types, on peut imaginer que la compagnie d'assurance classe chaque sinistre en type 1, 2, ... avec les probabilités $p_1, p_2, \ldots$ (les p_i étant positifs et de somme 1), ceci indépendamment des autres sinistres. Pour chaque $k \geq 1$, désignons par $N_k(t)$ le nombre de sinistres de type k durant $[0, t]$. Si, pour tout $t > 0$, on suppose que le nombre total de sinistres dans l'intervalle de temps $[0, t]$ est une variable aléatoire de Poisson de paramètre λt, on peut utiliser le résultat du Théorème 4.1.3.

(4) Précisons le classement des sinistres de la façon suivante : on dit qu'un sinistre est de type k, si son coût est égal à k unités de monnaie. Alors

$$N(t) = \sum_{k \geq 1} N_k(t) \qquad \text{est le nombre total de sinistres durant } [0, t];$$

$$S(t) = \sum_{k \geq 1} k\, N_k(t) \qquad \text{est le coût total des sinistres durant } [0, t].$$

On peut imaginer une *autre expression* pour le coût total des sinistres : dans l'intervalle $[0,t]$ se produisent $N(t)$ sinistres, qui coûtent X_1, X_2, …, $X_{N(t)}$ unités de monnaie, respectivement. Le coût total de ces sinistres est donc $S'(t) = X_1 + \cdots + X_{N(t)}$. Supposons que les coûts X_i sont des variables aléatoires, équidistribuées, de loi $\sum_k p_k \varepsilon_k$ et que $(N(t), X_1, X_2, \ldots)$ est une suite de variables aléatoires *indépendantes*, la variable $N(t)$ suivant la loi de Poisson de paramètre λt. Chaque variable X_i est ainsi interprétée comme le coût d'un sinistre non personnalisé. Sous ces hypothèses, on a le résultat remarquable suivant.

Théorème 4.1.4. *Les deux sommes $S(t)$ et $S'(t)$ ont même loi de probabilité. L'expression de $S'(t)$ montre que cette loi est une loi de Poisson composée.*

Démonstration. Notons $G_X(u) := \sum_{k\geq 1} p_k u^k$ la fonction génératrice de chaque X_i. Par ailleurs, les variables $N(t)$ et $N_k(t)$ ont pour fonctions génératrices respectives : $G_{N(t)}(u) = e^{\lambda t(u-1)}$, $G_{N_k(t)}(u) = e^{\lambda p_k t(u-1)}$.

Comme $G_{kN_k(t)}(u) = \mathbb{E}[u^{kN_k(t)}] = G_{N_k(t)}(u^k)$, la fonction génératrice de $S(t)$ est égale à

$$\begin{aligned} G_{S(t)}(u) &= \prod_{k\geq 1} G_{kN_k(t)}(u) = \prod_{k\geq 1} G_{N_k(t)}(u^k) = \exp(\lambda t \sum_{k\geq 1} p_k(u^k - 1)) \\ &= \exp(\lambda t(\sum_{k\geq 1} p_k u^k - 1)) = \exp(\lambda t(G_X(u) - 1)) \\ &= G_{N(t)} \circ G_X(u), \end{aligned}$$

qui est précisément la fonction génératrice de $S'(t)$. □

4.2 PROCESSUS DE POISSON MARQUÉS NON HOMOGÈNES

On reprend les mêmes notations que dans le paragraphe précédent, mais on étudie le cas général où la fonction $g(s)$ n'est *plus constante*. On suppose, toutefois, que r est égal à 2, de sorte que, pour tout $s \geq 0$, la loi de probabilité $g(s)$ est complètement déterminée par $g_1(s)$. Ainsi $g_1(s)$ (resp. $1 - g_1(s)$) est la probabilité qu'un top se produisant à l'instant s soit classé en type 1 (resp. en type 2). On suppose, en outre, que la fonction $g_1 : \mathbb{R}^+ \to [0,1]$ est intégrable sur tout intervalle borné.

D'après le Théorème 3.5.1, on sait que conditionnellement à l'évènement $\{N(t) = n\}$, les instants S_1, … , S_n où se produisent les n tops sont indépendants et uniformément répartis sur $[0,t]$. Pour décrire que le classement d'un top en type 1 ou 2 se fait *indépendamment de tous les autres tops*, on introduit une suite $(Y_1, \ldots, Y_n)$ de variables aléatoires toutes uniformément réparties sur $[0,1]$ et on fait l'hypothèse que, conditionnellement à $\{N(t) = n\}$, la suite $(S_1, Y_1, \ldots, S_n, Y_n)$ est formée de variables indépendantes. Ensuite, on donne au top se produisant en S_i le type 1, si et seulement si $Y_i \leq g_1 \circ S_i$.

Montrons que l'on réalise ainsi le classement voulu, c'est-à-dire,

$$\mathrm{P}\{Y_i \le g_1 \circ S_i \,|\, S_i = s\} = \mathbb{E}[I_{\{Y_i \le g_1 \circ S_i\}} \,|\, S_i = s] = g_1(s). \tag{4.2.1}$$

En effet,

$$\mathbb{E}[I_{\{Y_i \le g_1 \circ S_i\}} \,|\, S_i = s] = \int_{\mathbb{R}} I_{\{y \le g_1(s)\}}\, f_{Y_i \,|\, S_i}(y \,|\, s)\, dy,$$

où $f_{Y_i \,|\, S_i}(y \,|\, s)$ est la densité de Y_i liée par $\{S_i = s\}$, qui est la même que la densité de Y_i, à savoir la densité uniforme sur $[0,1]$, puisqu'on a supposé S_i et Y_i indépendantes. De là,

$$\mathbb{E}[I_{\{Y_i \le g_1 \circ S_i\}} \,|\, S_i = s] = \int_0^1 I_{\{y \le g_1(s)\}}\, dy = \int_0^{g_1(s)} dy = g_1(s). \quad \square$$

Théorème 4.2.1. *Soit $t > 0$ fixé ; conditionnellement à l'évènement $\{N(t) = n\}$, la variable aléatoire $N_1(t)$ suit une loi binomiale de paramètre (n, p), où*

$$p = \frac{1}{t} \int_0^t g_1(s)\, ds. \tag{4.2.2}$$

De plus, pour tout $t > 0$, les variables aléatoires $N_1(t)$ et $N_2(t)$ sont indépendantes et suivent une loi de Poisson de paramètres respectifs $\lambda t p$ et $\lambda t(1-p)$.

Démonstration. En effet,

$$\mathbb{E}[I_{\{Y_i \le g_1 \circ S_i\}}] = \int_{\mathbb{R}} \mathbb{E}[I_{\{Y_i \le g_1 \circ S_i\}} \,|\, S_i = s]\, f_{S_i}(s)\, ds,$$

où $f_{S_i}(s)$ est la densité de S_i conditionnellement à $\{N(t) = n\}$, c'est-à-dire la densité de la loi uniforme sur $[0, t]$. De là, d'après (4.2.1),

$$\mathbb{E}[I_{\{Y_i \le g_1 \circ S_i\}}] = \int_{\mathbb{R}} \mathbb{E}[I_{\{Y_i \le g_1 \circ S_i\}} \,|\, S_i = s]\, \frac{1}{t}\, I_{[0,t]}(s)\, ds = \frac{1}{t} \int_0^t g_1(s)\, ds.$$

Or, conditionnellement à $\{N(t) = n\}$, la variable aléatoire $N_1(t)$ a même loi que la somme des n variables de Bernoulli, indépendantes :

$$I_{\{Y_1 \le g_1 \circ S_1\}} + \cdots + I_{\{Y_n \le g_1 \circ S_n\}}.$$

C'est donc bien une variable aléatoire *binomiale* de paramètre (n, p). On refait alors *le même* calcul que dans la démonstration du Théorème 4.1.2, pour conclure. $\square$

Application 1. On suppose que des clients arrivent dans un magasin selon un processus de Poisson de densité $\lambda > 0$, que, dès leur arrivée, les clients sont servis immédiatement, qu'enfin les temps de service sont des variables aléatoires *indépendantes* admettant toutes la même fonction de répartition F. On demande

(a) la loi de probabilité du nombre de clients servis jusqu'à l'instant t (inclus) ;

(b) la loi de probabilité du nombre de clients dont le service n'est pas achevé à l'instant t.

On peut proposer la solution suivante : convenons d'appeler un client de type 1 ou de type 2, selon que son service est achevé ou non à l'instant t et désignons par $N_1(t)$, $N_2(t)$, le nombre de clients de type 1, 2, arrivés dans le magasin dans l'intervalle $[0, t]$.

Un client arrivant à l'instant $s \in [0, t]$ est de type 1, si son temps de service est au plus égal à $t - s$; la probabilité d'un tel évènement est $\mathrm{F}(t - s)$. Un client arrivant à l'instant $s \in [0, t]$ est de type 2, si son temps de service est supérieur à $t - s$; la probabilité d'un tel évènement est $1 - \mathrm{F}(t - s)$.

D'après le théorème précédent, le nombre $N_1(t)$ de clients de type 1 servis à l'instant t est une *variable aléatoire de Poisson de paramètre* λtp avec

$$p = \frac{1}{t}\int_0^t \mathrm{F}(t-s)\,ds = \frac{1}{t}\int_0^t \mathrm{F}(u)\,du,$$

et l'on a :

$$\mathbb{E}[N_1(t)] = \lambda tp = \lambda \int_0^t \mathrm{F}(u)\,du.$$

De même, $N_2(t)$ est une variable aléatoire de Poisson de paramètre $\lambda t(1-p)$ et en outre $N_1(t)$ et $N_2(t)$ sont indépendants.

Application 2. On suppose que des immigrants arrivent dans un camp de transit selon un processus de Poisson de densité $\lambda > 0$, que le temps de séjour d'un immigrant dans le camp est une variable aléatoire de fonction de répartition F et que ces temps de séjour sont des variables aléatoires *indépendantes*. On désigne par $N_1(t)$ le nombre d'immigrants qui ont quitté le camp avant l'instant t et par $N_2(t)$ le nombre d'immigrants présents dans le camp à l'instant t. Alors $N_1(t)$ et $N_2(t)$ sont des variables aléatoires dont les lois de probabilité sont données par l'Application 1.

4.3 SOMME DE DEUX PROCESSUS DE POISSON

Définition 4.3.1. Étant donnés deux processus de Poisson *indépendants* $\{N_1(t) : t \geq 0\}$ et $\{N_2(t) : t \geq 0\}$, de densités respectives λ_1, λ_2, on appelle *somme de ces deux processus*, le processus $\{N(t) : t \geq 0\}$, défini par $N(t) = N_1(t) + N_2(t)$ pour tout $t \geq 0$.

Proposition 4.3.1. *Le processus-somme* $\{N(t) : t \geq 0\}$ *est un processus de Poisson de densité* $\lambda_1 + \lambda_2$.

Démonstration. En effet, les variables $N_1(t)$ et $N_2(t)$ sont des variables aléatoires de Poisson, indépendantes, de paramètres respectifs $\lambda_1 t$ et $\lambda_2 t$. D'où $N(t)$ est une variable aléatoire de Poisson de paramètre $(\lambda_1 + \lambda_2)t$. □

Remarque 1. Si $T_1^{(1)}$ et $T_2^{(1)}$ désignent les temps d'attente du premier top dans chacun des processus, le temps d'attente du premier top dans le processus-somme est $T^{(1)} = \min(T_1^{(1)}, T_2^{(1)})$. Comme les variables $T_1^{(1)}$ et $T_2^{(1)}$ sont indépendantes, exponentielles de paramètres respectifs λ_1 et λ_2, le temps

d'attente $T^{(1)}$ suit une loi exponentielle de paramètre $(\lambda_1 + \lambda_2)$, comme on peut le vérifier directement.

Remarque 2. La probabilité pour que le premier top du processus-somme soit un top du processus $N_1(t)$ est donné par $\mathrm{P}\{T_1^{(1)} < T_2^{(1)}\}$, qui vaut $\lambda_1/(\lambda_1+\lambda_2)$, compte-tenu de la remarque précédente (*cf.* [FF1], chap. 12, § 4, Exemple 2). On peut donc construire le processus $N_1(t)$ à partir du processus $N(t)$ par le procédé du théorème précédent, en prenant $p = \lambda_1/(\lambda_1 + \lambda_2)$.

4.4 PROCESSUS DE POISSON COMPOSÉS

Définition 4.4.1. Soit (X_n) $(n \geq 1)$ une suite de variables aléatoires, indépendantes, identiquement distribuées ; soit $\{N(t) : t \geq 0\}$ un processus de Poisson, de densité $\lambda > 0$, indépendant de la suite (X_n) $(n \geq 1)$. Pour tout $t \geq 0$, on pose :

$$S(t) = \sum_{k=1}^{N(t)} X_k. \tag{4.4.1}$$

Le processus $\{S(t) : t \geq 0\}$ est alors appelé *processus de Poisson composé.*

Exemple. On suppose que des clients entrent dans un grand magasin selon un processus de Poisson $\{N(t) : t \geq 0\}$ de densité λ. On suppose, de plus, que les montants X_n $(n \geq 1)$ des dépenses faites par chaque client forment une suite de variables aléatoires indépendantes et identiquement distribuées. On suppose, en outre, que la suite (X_n) $(n \geq 1)$ est indépendante du processus $\{N(t) : t \geq 0\}$. Avec les notations précédentes, $S(t)$ est le montant des dépenses faites dans le magasin dans l'intervalle $[0, t]$.

Le processus de Poisson composé $S(t)$ avait déjà été rencontré dans le Théorème 4.1.4, où il avait été noté $S'(t)$ et où les variables X_i étaient supposées à valeurs dans $\mathbb{N}$. Pour obtenir une information sur la loi de $S(t)$, lorsque les variables X_i sont quelconques, on détermine la *fonction caractéristique* de $S(t)$, au lieu de sa fonction génératrice.

Théorème 4.4.1. *Soit $\varphi(u)$ la fonction caractéristique commune des X_n, soit $G_{N(t)}(u)$ la fonction génératrice de $N(t)$. Alors la fonction caractéristique de $S(t)$ est donnée par :*

$$\varphi_{S(t)}(u) = G_{N(t)} \circ \varphi(u) = e^{\lambda t(\varphi(u)-1)}.$$

Démonstration. En effet,

$$\begin{aligned}\varphi_{S(t)}(u) &= \mathbb{E}[e^{iuS(t)}] = \mathbb{E}[\mathbb{E}[e^{iuS(t)} \mid N(t)]] \\ &= \sum_{n\geq 0} \mathbb{E}[e^{iuS(t)} \mid N(t) = n]\mathrm{P}\{N(t) = n\}\end{aligned}$$

$$= \sum_{n\geq 0} \mathbb{E}[e^{iu(X_1+\cdots+X_{N(t)})} \mid N(t) = n]\mathrm{P}\{N(t) = n\}$$
$$= \sum_{n\geq 0} \mathbb{E}[e^{iu(X_1+\cdots+X_n)} \mid N(t) = n]\mathrm{P}\{N(t) = n\}.$$

D'où, puisque $N(t)$ est indépendant de $(X_1, \ldots, X_n)$,

$$\varphi_{S(t)}(u) = \sum_{n\geq 0} \mathbb{E}[e^{iu(X_1+\cdots+X_n)}]\mathrm{P}\{N(t) = n\}$$
$$= \sum_{n\geq 0} (\varphi(u))^n \mathrm{P}\{N(t) = n\} = G_{N(t)}(\varphi(u)) = G_{N(t)} \circ \varphi(u). \quad \square$$

En dérivant, on en tire : $\mathbb{E}[S(t)] = \lambda t\, \mathbb{E}[X]$, $\quad \mathrm{Var}\, S(t) = \lambda t\, \mathbb{E}[X^2]$.

Remarque. Chacun des processus $\{N_1(t) : t \geq 0\}$, $\{N_2(t) : t \geq 0\}$, décrit dans le Théorème 4.2.1, peut être interprété comme un *processus de Poisson composé*. Montrons-le pour $N_1(t) := \sum_{k=1}^{N(t)} I_k$, où les I_k sont des variables de Bernoulli, indépendantes, de loi commune $q\varepsilon_0 + p\varepsilon_1$, elles-mêmes indépendantes du processus $\{N_1(t) : t \geq 0\}$.

En effet, conditionnellement à $\{N(t) = n\}$, la variable aléatoire $N_1(t)$ est binomiale de paramètres (n, p). La fonction génératrice $G_{N_1(t)}(u)$ de $N_1(t)$ vaut donc :

$$G_{N_1(t)}(u) = G_{N(t)}(u) \circ G_{I_1}(u) = e^{-\lambda t(pu+1-p-1)} = e^{\lambda pt(u-1)}. \quad \square$$

4.5 PROCESSUS DE POISSON ET FONCTIONS GAMMA ET BÊTA

On rappelle (*cf.* [FF1], chap. 15, § 3, Exemple 3) que si (X, Y) est un couple de variables aléatoires de densité conjointe $f(x, y)$, la variable aléatoire

$$U = \begin{cases} \dfrac{X}{Y}, & \text{si } Y \neq 0\,; \\ 0, & \text{si } Y = 0\,; \end{cases}$$

a une densité de probabilité donnée par

$$h(u) = \int_{-\infty}^{+\infty} |v|\, f(uv, v)\, dv \quad (u \in \mathbb{R}).$$

On reprend les notations utilisées dans ce chapitre et le précédent pour un processus de Poisson de densité $\lambda > 0$: les S_n $(n \geq 1)$ sont les instants ou se produisent les tops, $T_1 = S_1$ est le temps d'attente du premier top et T_n $(n \geq 2)$ est le temps séparant le $(n-1)^{\text{ième}}$ top du $n^{\text{ième}}$. On pose $S_0 = 0$ et l'on a : $S_n = T_1 + \cdots + T_n$ $(n \geq 1)$. On a déjà vérifié que S_n suivait une *loi gamma de paramètre* (n, λ) (*cf.* Théorème 3.2.2).

Considérons le rapport $U(r,s) = \dfrac{S_r}{S_{r+s}}$ $(r, s \geq 0)$. Comme $0 \leq S_r \leq S_{r+s}$, on a $0 \leq U(r,s) \leq 1$; de plus, $U(r,s)$ est le rapport d'une variable aléatoire de loi $\Gamma(r,\lambda)$ et d'une variable aléatoire de loi $\Gamma(r+s,\lambda)$. Ces deux variables aléatoires ne sont pas indépendantes. Pourtant la variable aléatoire $U(r,s)$ a une loi de probabilité qu'on peut calculer et qui *ne dépend pas de* λ, comme le montre le théorème suivant.

Théorème 4.5.1. *La variable* $U(r,s) = \dfrac{S_r}{S_{r+s}}$ $(r, s \geq 0)$ *suit la loi bêta de première espèce* $B_1(r,s)$*, dont la densité est donnée par :*

$$\frac{1}{B(r,s)} u^{r-1}(1-u)^{s-1} \qquad (0 \leq u \leq 1).$$

(*cf.* [FF1], chap. 14, § 9.)

Démonstration. Posons $S_{r+s} = S_r + S'_s$ où $S'_s = T_{r+1} + \cdots + T_{r+s}$, de sorte que les variables aléatoires S_r et S'_s sont indépendantes, de lois respectives $\Gamma(r,\lambda)$ et $\Gamma(s,\lambda)$ et de densités respectives $f(r,\lambda,x)$ et $f(s,\lambda,y)$. La densité du couple (S_r, S'_s) est donc $f(r,\lambda,x)f(s,\lambda,y)$ sur $\mathbb{R}_+ \times \mathbb{R}_+$. Posons $u = x$ et $v = x + y$, de sorte que $(S_r, S_{r+s}) = (u \circ (S_r, S'_s), v \circ (S_r, S'_s))$. La densité du couple (S_r, S_{r+s}) vaut donc

$$\begin{aligned}
&f(r,\lambda,x(u,v))f(s,\lambda,y(u,v)) \left| \frac{D(x,y)}{D(u,v)} \right| \\
&\quad = f(r,\lambda,u)f(s,\lambda,v-u) \begin{vmatrix} 1 & 0 \\ -1 & 1 \end{vmatrix} = f(r,\lambda,u)f(s,\lambda,v-u) \\
&\quad = \lambda e^{-\lambda u} \frac{(\lambda u)^{r-1}}{(r-1)!} \lambda e^{-\lambda(v-u)} \frac{(\lambda(v-u))^{s-1}}{(s-1)!} = \lambda^{r+s} e^{-\lambda v} \frac{u^{r-1}(v-u)^{s-1}}{(r-1)!\,(s-1)!},
\end{aligned}$$

si $0 \leq u \leq v$ et 0 autrement.

La densité $h(u)$ de $U(r,s)$ vaut donc 0 si $u \notin [0,1]$ et lorsque $u \in [0,1]$

$$\begin{aligned}
h(u) &= \int_0^{+\infty} v\, \lambda^{r+s} e^{-\lambda v} \frac{(uv)^{r-1}(v-vu)^{s-1}}{(r-1)!\,(s-1)!}\, dv \\
&= \lambda^{r+s} \frac{u^{r-1}(1-u)^{s-1}}{(r-1)!\,(s-1)!} \int_0^{+\infty} e^{-\lambda v} v^{r+s-1}\, dv \\
&= \frac{u^{r-1}(1-u)^{s-1}}{(r-1)!\,(s-1)!} \int_0^{+\infty} e^{-t} t^{r+s-1}\, dv \qquad [\text{avec } \lambda v = t] \\
&= \frac{u^{r-1}(1-u)^{s-1}}{\Gamma(r)\,\Gamma(s)} \Gamma(r+s) = \frac{u^{r-1}(1-u)^{s-1}}{B(r,s)}. \quad \square
\end{aligned}$$

Considérons maintenant deux processus de Poisson, indépendants, respectivement de densités λ et μ $(\lambda, \mu > 0)$, que l'on étudie à partir de l'instant 0. Les instants où se produisent les tops dans le processus de densité μ sont notés $S'_1, S'_2, \ldots$ On pose, de plus, $T'_n = S'_n - S'_{n-1}$ $(n \geq 1)$ avec $S'_0 = 0$. Les

deux processus étant indépendants, les suites (S_n) et (S'_n) sont elles aussi indépendantes. Dans le théorème suivant, on détermine la loi du rapport $V(r,s) = \dfrac{S_r}{S'_s}$ $(r, s \geq 0)$.

Théorème 4.5.2. *La variable aléatoire* $V(r,s) = \dfrac{S_r}{S'_s}$ $(r, s \geq 0)$ *suit une loi bêta de seconde espèce, dont la densité est donnée par*

$$h(v) = \begin{cases} \dfrac{1}{B(r,s)} \dfrac{\lambda^r \mu^s v^{r-1}}{(\lambda v + \mu)^{r+s}}, & \textit{si } v \geq 0\,; \\ 0, & \textit{ailleurs.} \end{cases}$$

Démonstration. La densité $f(x,y)$ du couple (S_r, S'_s) est donnée par le produit des densités, c'est-à-dire, par $e^{-(\lambda x + \mu y)} \dfrac{\lambda^r \mu^s x^{r-1} y^{s-1}}{(r-1)!\,(s-1)!}$ pour x et y positifs et 0 ailleurs.

On a donc $h(v) = 0$ pour $v < 0$; pour $v \geq 0$, on obtient :

$$\begin{aligned} h(v) &= \int_0^{+\infty} u\, f(uv, u)\, du \\ &= \int_0^{+\infty} \frac{\lambda^r \mu^s v^{r-1}}{(r-1)!\,(s-1)!} \int_0^{+\infty} e^{-u(\lambda v + \mu)} u^{r+s-1}\, du \\ &= \frac{\lambda^r \mu^s}{(r-1)!\,(s-1)!} \frac{v^{r-1}}{(\lambda v + \mu)^{r+s}} \int_0^{+\infty} e^{-t} t^{r+s-1}\, dt \qquad [u(\lambda v + \mu) = t] \\ &= \frac{1}{B(r,s)} \frac{\lambda^r \mu^s v^{r-1}}{(\lambda v + \mu)^{r+s}}. \quad \square \end{aligned}$$

4.6 ABANDON DE L'HYPOTHÈSE DES ACCROISSEMENTS STATIONNAIRES

On étudie maintenant un modèle de processus, où la densité λ dépend du temps. Les axiomes qui avaient été énoncés pour les processus de Poisson à accroissements stationnaires sont remplacés par les suivants.

Définition 4.6.1. Soit λ une fonction réelle continue définie sur $\mathbb{R}_+$. Le processus de comptage $\{N(t) : t \geq 0\}$ est appelé *processus de Poisson de fonction de densité* $\lambda(t)$, s'il satisfait les axiomes suivants.

(a) $N(0) = 0$;

(b) Le processus est à accroissements indépendants.

(c) Soit $m(t) = \int_0^t \lambda(u)\, du$; alors pour tout couple (s,t) tel que $0 \leq s < t < +\infty$, le nombre $N(t) - N(s)$ de tops situés dans $]s,t]$ est une variable aléatoire de Poisson de paramètre

$$\mu(\,]s,t]\,) = m(t) - m(s) = \int_s^t \lambda(u)\, du.$$

Comme $N(t)$ est une variable de Poisson de paramètre $m(t)$, on a :

$$m(t) = \mathbb{E}[N(t)], \qquad m(t) = \operatorname{Var} N(t).$$

Pour cette raison, $m(t)$ est appelée la *fonction moyenne du processus* ou *fonction de renouvellement*. C'est une fonction définie pour $t \geq 0$, à valeurs positives, croissante, continue et dérivable pour $t \geq 0$, sa dérivée étant $\lambda(t)$. En outre, $m(0) = 0$. On suppose que $m(t)$ tend vers l'infini avec t.

Notons encore les ordres de grandeur suivants, pour $h > 0$:

$$\begin{aligned}
\mathrm{P}\{N(t+h) - N(t) = 1\} &= (m(t+h) - m(t))e^{-(m(t+h)-m(t))} \\
&= \lambda(t)h + o(h)\,; \\
\mathrm{P}\{N(t+h) - N(t) \geq 2\} &= 1 - e^{-(m(t+h)-m(t))}(1 + (m(t+h) - m(t)) \\
&= \frac{\lambda^2(t)h^2}{2} + o(h^3) = o(h).
\end{aligned}$$

Proposition 4.6.1. *Tout processus de Poisson, non à accroissements stationnaires, peut se ramener à un processus à accroissements stationnaires par un changement d'horloge.*

Démonstration. Soit $\{N(t) : t \geq 0\}$ un processus de Poisson, non à accroissements stationnaires, de fonction de densité $\lambda(t)$, supposée *strictement positive* pour tout $t \geq 0$. La fonction de renouvellement $m(t) = \int_0^t \lambda(u)\,du$ correspondante est alors une fonction continue, strictement croissante, nulle pour $t = 0$ et qui tend vers l'infini lorsque t tend vers l'infini. Elle admet donc une fonction inverse m^{-1}. Posons $N^*(t) = N(m^{-1}(t))$. Le processus de comptage $\{N^*(t) : t \geq 0\}$ est un processus de Poisson, à accroissements stationnaires, de densité $\lambda = 1$, puisque $N^*(t)$ est une variable aléatoire de Poisson de paramètre $m(m^{-1}(t)) = t$. □

Donnons quelques propriétés des processus de Poisson non à accroissements stationnaires.

Propriété 4.6.2. *Soit t un point de discontinuité du processus. Avec une probabilité égale à 1, l'amplitude du saut en t est égale à 1.*

Démonstration. Soit $\varepsilon > 0$ et soit I_ε l'intervalle $]t - \varepsilon, t + \varepsilon]$.En notant N_ε le nombre de tops situés dans I_ε, on a :

$$\begin{aligned}
\mathrm{P}\{N_\varepsilon > 1 \mid N_\varepsilon \geq 1\} &= \frac{\mathrm{P}\{N_\varepsilon > 1\}}{\mathrm{P}\{N_\varepsilon \geq 1\}} \\
&= \frac{1 - e^{-(m(t+\varepsilon)-m(t-\varepsilon))}(1 + m(t+\varepsilon) - m(t-\varepsilon))}{1 - e^{-(m(t+\varepsilon)-m(t-\varepsilon))}} \\
&= 2\varepsilon\,\lambda(t) + o(\varepsilon),
\end{aligned}$$

qui tend vers 0, lorsque ε tend vers 0. □

Propriété 4.6.3. *Pour $0 \leq s < t < +\infty$, la probabilité qu'il se produise une infinité de tops dans $]s,t]$ est nulle, ou encore, avec probabilité égale à 1, la suite des points de discontinuité du processus n'admet pas de point d'adhérence à distance finie.*

Démonstration. En effet,

$$\mathrm{P}\{N(t) - N(s) < +\infty\} = \sum_{n \geq 0} e^{-(m(t)-m(s))} \frac{(m(t) - m(s))^n}{n!} = 1. \quad \square$$

Propriété 4.6.4. *La fonction aléatoire $N(t)$ est localement continue en probabilité en tout point $t > 0$.*

Démonstration. Soient $t > 0$ et $\varepsilon > 0$ tels que $t - \varepsilon > 0$. Alors

$$\mathrm{P}\{N(t+\varepsilon) - N(t-\varepsilon) > 0\} = 1 - e^{m(t+\varepsilon)-m(t-\varepsilon)} \sim 2\lambda(t)\varepsilon,$$

qui tend vers 0 avec ε. $\square$

Théorème 4.6.5. *Soit $0 \leq s < t < +\infty$. Conditionnellement lorsque $N(t) - N(s) = n$, les n instants tops situés dans $]s,t]$ ont même loi de probabilité qu'un n-uple de variables aléatoires indépendantes identiquement distribuées selon la loi de probabilité de fonction de répartition $F(u) = \dfrac{m(u) - m(s)}{m(t) - m(s)}$ $(s < u \leq t)$. En d'autres termes, pour $0 < s < s' < t' < t < +\infty$, on a :*

$$\begin{aligned}\mathrm{P}\{N(t') - N(s') = k \,|\, N(t) - N(s) = n\}& \\ &= \binom{n}{k} \Big(\frac{m(t') - m(s')}{m(t) - m(s)}\Big)^k \Big(1 - \frac{m(t') - m(s')}{m(t) - m(s)}\Big)^{n-k},\end{aligned}$$

pour $k = 0, 1, \ldots, n$.

La démonstration est analogue à la démonstration qui avait été donnée pour les processus à accroissements stationnaires.

Donnons maintenant la construction d'un processus de Poisson, non à accroissements stationnaires, de fonction de densité $\lambda(t)$. Si cette fonction est *bornée*, ce processus peut être construit à partir d'un processus de Poisson, à accroissements stationnaires. Soit $\lambda > 0$ tel que pour tout $t \geq 0$ on ait $\lambda(t) \leq \lambda$ et soit $\{N(t) : t \geq 0\}$ un processus de Poisson à accroissements stationnaires de densité λ. On convient de ce qui suit : si un top de ce processus se produit à l'instant s, on le retient avec la probabilité $g(s) = \lambda(s)/\lambda$ et on le rejette avec la probabilité $1 - g(s)$. On désigne par $\{N_1(t) : t \geq 0\}$ le processus de comptage défini par les tops retenus.

Proposition 4.6.6. *Le processus $\{N_1(t) : t \geq 0\}$ est un processus de Poisson, non à accroissements stationnaires, dont la fonction de densité est $\lambda(t)$.*

Démonstration. Ce théorème a, en fait, déjà été démontré, la variable de Poisson $N_1(t)$ étant de paramètre $\lambda p t = m(t)$, puisque p est donné par :

$$p = \frac{1}{t} \int_0^t g(s)\, ds = \frac{1}{t\lambda} \int_0^t \lambda(s)\, ds = \frac{1}{\lambda t} m(t). \quad \square$$

Nous donnons une seconde définition des processus de Poisson non à accroissements stationnaires.

Définition 4.6.2. Le processus de comptage $\{N(t) : t \geq 0\}$ est appelé *processus de Poisson de fonction de densité* $\lambda(t)$, s'il vérifie :

a) $N(0) = 0$;

b) le processus est à accroissements indépendants;

c) pour $h \downarrow 0$, on a : $\mathrm{P}\{N(t+h) - N(t) = 1\} = \lambda(t) + o(h)$ et $\mathrm{P}\{N(t+h) - N(t) \geq 2\} = o(h)$.

Application. Soit $(X_1, X_2, \dots)$ une suite de variables aléatoires, indépendantes, identiquement distribuées, à valeurs positives, de densité commune f, de fonction de répartition commune F. On pose $\lambda(t) = \dfrac{f(t)}{1 - \mathrm{F}(t)}$ (taux de défaillance commune). On convient d'enregistrer la valeur de la $n^{\text{ième}}$ variable aléatoire X_n, si $X_n > \max(X_1, \dots, X_{n-1})$ $(X_0 = 0)$ et on désigne par $N(t)$ le nombre de valeurs enregistrées et situées dans $[0, t]$. Dans l'exemple ci-après, $N(t)(\omega) = 5$.

$$0 \quad x_1 \quad x_3 \quad x_2\, x_6 \quad x_4 \qquad x_5 \quad x_7 \qquad t \qquad x_8$$

Proposition 4.6.7. *Le processus de comptage* $\{N(t) : t \geq 0\}$ *est un processus de Poisson non à accroissements stationnaires, de fonction de densité* $\lambda(t)$.

Démonstration. L'évènement $\{N(t+h) - N(t) \geq 1\}$ se réalise, si et seulement s'il existe un entier $n \geq 1$ tel que les évènements $\{X_n \in]t, t+h]\}$ et $\{X_1 \leq t, \dots, X_{n-1} \leq t\}$ se réalisent. Par conséquent,

$$\begin{aligned}
\mathrm{P}\{N(t+h) - N(t) \geq 1\} &= \mathrm{P}\{\bigcup_{n\geq 1} \{X_1 \leq t, \dots, X_{n-1} \leq t\} \cap \{X_n \in]t, t+h]\}\} \\
&= \sum_{n\geq 1} \mathrm{P}\{X_1 \leq t, \dots, X_{n-1} \leq t\} \mathrm{P}\{X_n \in]t, t+h]\} \\
&= \sum_{n\geq 1} \mathrm{F}^{n-1}(t)(\mathrm{F}(t+h) - \mathrm{F}(t)) \\
&= \frac{\mathrm{F}(t+h) - \mathrm{F}(t)}{1 - \mathrm{F}(t)}.
\end{aligned}$$

Or, ce dernier terme est équivalent à

$$\frac{h\, f(t)}{1 - \mathrm{F}(t)} + o(h) = h\, \lambda(t) + o(h). \quad \square$$

Cas particulier. Supposons que la loi commune des X_n soit la loi exponentielle de paramètre $\lambda > 0$. Alors $\lambda(t) = f(t)/(1 - \mathrm{F}(t)) = \lambda e^{-\lambda t}/e^{-\lambda t} = \lambda$. Le processus $\{N(t) : t \geq 0\}$ est alors un processus de Poisson à accroissements stationnaires, de densité λ. Le procédé précédent, dit des *valeurs saillantes*,

fournit donc un procédé pour construire un processus à accroissements stationnaires à partir d'une suite (X_n) de variables aléatoires indépendantes, toutes de loi exponentielle de paramètre λ.

4.7 ESTIMATION DE LA DENSITÉ D'UN PROCESSUS DE POISSON

Soit un processus de Poisson $\{N(t) : t \geq 0\}$, de densité λ *inconnue*. On se propose d'estimer cette densité à partir d'observations. Donnons deux méthodes.

1) Supposons qu'on observe le processus jusqu'à un instant $T > 0$. On dispose alors des données suivantes :

a) n, le nombre de tops dans $[0, T]$;

b) les instants $x_1, \ldots, x_n$ $(0 < x_1 < x_2 < \cdots < x_n < T)$, où se sont produits les n tops consécutifs dans $[0, T]$.

La vraisemblance de ces observations est donnée par

$$L(n, x_1, \ldots, x_n; \lambda) = \mathrm{P}\{N(T) = n\}\, g(x_1, \ldots, x_n \,|\, N(T) = n)$$
$$= e^{-\lambda T}\frac{(\lambda T)^n}{n!}\frac{n!}{T^n} = \lambda^n e^{-\lambda T} \quad (0 < x_1 < \cdots < x_n < T),$$

d'où

$$\operatorname{Log} L = n \operatorname{Log} \lambda - \lambda T$$

et

$$\frac{\partial}{\partial \lambda}\operatorname{Log} L = \frac{n}{\lambda} - T = 0 \quad \Rightarrow \quad \widehat{\lambda} = \frac{n}{T}.$$

Ainsi l'*estimation par le maximum de vraisemblance* est donné par $\widehat{\lambda} = n/T$. L'*estimateur* correspondant est obtenu en remplaçant n par $N(T)$: $\widehat{\lambda} = \dfrac{N(T)}{T}$.

Proposition 4.7.1. *L'estimateur $\widehat{\lambda}$ est non biaisé, exhaustif et complet.*

Démonstration. En effet, $\mathbb{E}[\widehat{\lambda}] = (1/T)\,\mathbb{E}[N(T)] = (1/T)\,\lambda T = \lambda$. Comme la vraisemblance $L(x_1, \ldots, x_n; \lambda)$ est de la forme $\lambda^n e^{-\lambda T}$, l'estimateur est bien exhaustif. Enfin, comme $N(T)$ suit une loi de Poisson de paramètre λT et que la loi de Poisson est complète, l'estimateur $\widehat{\lambda}$ est complet. □

Puisque $\widehat{\lambda}$ est non-biaisé, exhaustif, complet, c'est l'unique estimateur non-biaisé de variance minimum de λ.

2) Supposons que nous observions le processus jusqu'à l'apparition du $n^{\text{ième}}$ top. Nous disposons alors de la donnée des instants $x_1, \ldots, x_n$ des apparitions des tops. La vraisemblance est donnée par :

$$L(x_1, \ldots, x_n; \lambda) = \lambda e^{-\lambda x_1} \ldots \lambda e^{-\lambda x_n} = \lambda^n e^{-\lambda(x_1 + \cdots + x_n)}$$
$$\operatorname{Log} L = n \operatorname{Log} \lambda - \lambda(x_1 + \cdots + x_n)\,;$$

par suite,

$$\frac{\partial}{\partial \lambda} \log L = \frac{n}{\lambda} - (x_1 + \cdots + x_n) = 0 \quad \Rightarrow \quad \widehat{\lambda} = \frac{n}{x_1 + \cdots + x_n}.$$

En posant $S_n = X_1 + \cdots + X_n$, instant d'arrivée du $n^{\text{ième}}$ top, on voit que l'estimateur correspondant est $\widehat{\lambda} = n/S_n$.

4.8 COMPLÉMENTS ET EXERCICES

4.1 À raison d'un véhicule toutes les dix secondes en moyenne, le flux des véhicules dans une voie donnée comporte une proportion de $p = 10$ % de camions et de 90 % de voitures (particulières).

(a) Quelle est la probabilité qu'au moins un camion passe dans un intervalle d'une minute ?

(b) Sachant que dix camions sont passés dans un intervalle de cinq minutes, quel est le nombre moyen de véhicules qui sont passés dans cet intervalle ?

(c) Trente véhicules sont passés durant dix minutes, quelle est la probabilité que trois parmi eux sont des camions ?

4.2 Des voyageurs transitent à un poste frontière de la Communauté européenne, les citoyens à un rythme de λ par unité de temps, ceux étrangers à la Communauté à un rythme de μ par unité de temps. À partir d'un instant donné, évaluer, en fonction de λ et μ, la probabilité que le $n^{\text{ième}}$ citoyen de la Communauté se présente avant le $m^{\text{ième}}$ étranger. [Utiliser l'identité

$$\sum_{k=0}^{r} e^{-\lambda} \frac{\lambda^k}{k!} = \int_{\lambda}^{+\infty} \frac{t^r e^{-t}}{r!}\, dt,$$

que l'on établit immédiatement par intégration par parties.]

4.3 On reprend les notations du paragraphe 4.3. Soit F la fonction de répartition de X_1 ; la puissance $n^{\text{ième}}$ du produit de convolution de F est notée F^{*n}. La fonction de répartition de $S(t)$ est donnée par :

$$\mathrm{P}\{S(t) \le x\} = \sum_{n=0}^{+\infty} \mathrm{F}^{*n}(x) \frac{1}{n!} e^{-\lambda t} (\lambda t)^n.$$

4.4 On reprend les notations du paragraphe 4.4. Soit (X_n) $(n \ge 1)$ une suite de variables aléatoires indépendantes, chacune de loi $q\varepsilon_{-1} + p\varepsilon_{+1}$. Pour $n \ge 1$, on pose $Y_n := X_1 + \cdots + X_n$ et $Y_0 := 0$.

(a) On a : $\mathrm{P}\{Y_{2n+m} = m\} = \binom{2n+m}{n} p^{n+m} q^n \quad (m \ge -n)$.

(b) Évaluer $\mathrm{P}\{S(t) = m\}$.

(c) Vérifier que $\sum_{n,m} \mathrm{P}\{S(t) = m, N(t) = n\} = 1$.

4.5 On considère un processus de Poisson $\{N(t) : t \geq 0\}$ de densité $\lambda > 0$. Étudier les processus suivants déduits du processus $\{N(t) : t \geq 0\}$ par les opérations suivantes.

(a) On supprime les tops de rang impair. Noter $\{N_1(t) : t \geq 0\}$ le processus obtenu.

(b) Pour tout $i = 1, 2, \ldots$ on décale l'instant S_i d'apparition du $i^{\text{ième}}$ top d'une quantité constante a. Noter $\{N_2(t) : t \geq 0\}$ le processus obtenu.

(c) Pour tout $i = 1, 2, \ldots$ on décale l'instant S_i d'apparition du $i^{\text{ième}}$ top d'une quantité aléatoire Y positive dont la densité de probabilité $g(x)$ est indépendante de i. [Soient $N_3(t)$ le processus résultant et $G(t)$ la fonction de répartition de Y. On démontre d'abord que, conditionnellement à l'évènement $\{N(t) = n\}$, la variable aléatoire $N_3(t)$ est binomiale de paramètre $(n, p(t))$ avec

$$p(t) = \frac{1}{t} \int_0^t G(t - x)\, dx,$$

puis on calcule $\mathrm{P}\{N_3(t) = k\}$.]

(d) Soient M_n le milieu du segment $[S_{n-1}, S_n]$ et V_n la longueur du segment $[M_{n-1}, M_n]$. Calculer les coefficients de corrélation des variables V_{n-1} et V_n, puis des variables V_{n-1} et V_{n+1}. Calculer la loi de probabilité de V_n.

4.6 (a) On considère un processus de Poisson $\{N(t) : t \geq 0\}$ de densité $\lambda > 0$. Pour tout $t \geq 0$ et tout entier $n \geq 0$, on pose $P_n(t) := \mathrm{P}\{N(t) = n\}$. Pour $t > 0$ et $n \geq 1$, soit $P_n'(t)$ la dérivée de $P_n(t)$. Alors, $P_n'(t) = -\lambda P_n(t) + \lambda P_{n-1}(t)$. De plus, $P_n(0) = 0$ pour $n \geq 1$. [Utiliser la Propriété 3.1.3.]

(b) Soit (λ_n) $(n \geq 0)$ une suite de nombres réels positifs. On considère maintenant un processus de comptage $\{N(t) : t \geq 0\}$, à accroissements indépendants. On suppose, de plus, que si l'évènement $\{N(t) = n\}$ est réalisé pour un certain $n \geq 0$, alors la probabilité qu'il se produise un top dans $]t, t+h]$ est égale à $\lambda_n h + o(h)$ et la probabilité qu'il se produise au moins deux tops est $o(h)$. Établir le système différentiel pour la fonction $P_n(t) := \mathrm{P}\{N(t) = n\}$:

$$\begin{cases} P_n'(t) = -\lambda_n\, P_n(t) + \lambda_{n-1}\, P_{n-1}(t) \quad (n = 1, 2, \ldots); \\ P_0'(t) = -\lambda_0\, P_0(t). \end{cases}$$

(c) Mêmes hypothèses que dans (b). On suppose, de plus, que pour tout $n \geq 0$, on a $\lambda_n = n\,\lambda$ et $N(0) = n_0$, où $\lambda > 0$ et $n_0 \geq 1$. Si $N(0) = n_0 = 1$, alors $P_0(t) = 0$ et $P_n(t) = e^{-\lambda t}(1 - e^{-\lambda t})^{n-1}$ pour $n \geq 1$. La fonction génératrice $G(s, t) := \mathbb{E}[s^{N(t)}]$ de $N(t)$ est égale à : $s\, e^{-\lambda t}(1 - s(1 - e^{-\lambda t}))^{-1}$. [Poser $Q_n(t) := P_n(t) e^{n\lambda t}$.]

Si $N(0) = n_0 \geq 2$, la fonction génératrice de $N(t)$ vaut $(G(s,t))^{n_0}$; d'où

$$P_n(t) = \begin{cases} \binom{n-1}{n-n_0} e^{-n_0\lambda t}(1-e^{-\lambda t})^{n-n_0}, & \text{si } n \geq n_0\,; \\ 0, & \text{si } n \leq n_0 - 1. \end{cases}$$

4.7 Processus de naissance et mort

On considère un processus de Poisson $\{N(t) : t \geq 0\}$, à accroissements stationnaires, de densité $\lambda > 0$. On interprète l'apparition d'un top comme la *naissance d'un individu*. On suppose, de plus, que chaque individu a une durée de vie X et que les durées de vie des différents individus sont des variables aléatoires *indépendantes, identiquement distribuées* et *indépendantes du processus*. Soient S_k la date de naissance du $k^{\text{ième}}$ individu et X_k sa durée de vie, de sorte que la date de sa mort sera $S_k + X_k$. On pose $S_0 := X_0 := 0$. On désigne par $Q(t)$ le nombre d'individus en vie à la date t. Il s'agit d'évaluer la loi de probabilité de $Q(t)$, dans l'hypothèse $N(0) = 0$.

(a) Soient $r(u) := \mathrm{P}\{X \geq u\}$ la fonction de survie de X et $\overline{r}(t)$ sa valeur moyenne $\overline{r}(t) := (1/t)\int_0^t r(u)\,du$. Si U est une variable aléatoire uniformément répartie sur $[0,t]$, indépendante de X, la fonction de survie de $U+X$ est donnée par : $\mathrm{P}\{U + X \geq t\} = \overline{r}(t)$.

(b) La fonction génératrice $h(u) := \mathbb{E}[u^{Q(t)}]$ de $Q(t)$ est donnée par $e^{-\lambda t\overline{r}(t)(1-u)}$.

(c) Sous l'hypothèse $N(0) = 0$, le nombre $Q(t)$ d'individus en vie à la date $t \geq 0$ est une variable aléatoire de Poisson de paramètre $m(t) = \lambda\int_0^t r(u)\,du$.

4.8 Processus de naissance et mort. Autre modèle

Soient (λ_n), (μ_n) $(n \geq 0)$ deux suites de nombres réels positifs. On considère maintenant un processus de comptage $\{N(t) : t \geq 0\}$, à accroissements indépendants. On suppose, de plus, que si l'évènement $\{N(t) = n\}$ est réalisé pour un certain $n \geq 0$, alors la probabilité qu'il se produise un *top positif* (ou une *naissance*) dans $]t, t+h]$ est égale à $\lambda_n h + o(h)$. Si $\{N(t) = n\}$ est réalisé pour un certain $n \geq 1$, la probabilité qu'il se produise un *top négatif* (ou une *mort*) est égale à $\mu_n h + o(h)$. Enfin, la probabilité qu'il se produise au moins deux tops (positifs et/ou négatifs) est $o(h)$. Le nombre $N(t)$ est égal au nombre de naissances diminué du nombre de morts dans l'intervalle $[0,t]$. On suppose que $N(0) = n_0 \geq 1$.

(a) En posant $P_n(t) := \mathrm{P}\{N(t) = n\}$, on a :

$$P_n(t+h) = P_n(t)(1 - \lambda_n h - \mu_n h) + \lambda_{n-1} h P_{n-1}(t) + \mu_{n+1} h P_{n+1}(t) + o(h), \quad \text{pour } n \geq 1\,;$$

$$P_0(t+h) = (1 - \lambda_0 h)P_0(t) + \mu_1 h P_1(t) + o(h).$$

(b) On en déduit :

$$P'_n(t) = -(\lambda_n + \mu_n)P_n(t) + \lambda_{n-1}P_{n-1}(t) + \mu_{n+1}P_{n+1}(t) \quad (n \geq 1)$$
$$P'_0(t) = -\lambda_0 P_0(t) + \mu_1 P_1(t)$$
$$P_n(0) = \delta_{n,n_0}.$$

(c) On prend le cas particulier $\lambda_n = n\lambda$ et $\mu_n = n\mu$ ($n \geq 0$), où $\lambda > 0$ et $\mu > 0$ et on désigne par $G(s,t) := \mathbb{E}[s^{N(t)}]$ la fonction génératrice de $N(t)$. Alors $G(s,t)$ satisfait l'équation aux dérivées partielles :

$$\frac{\partial G(s,t)}{\partial t} = (\lambda s - \mu)(s-1)\frac{\partial G(s,t)}{\partial s}, \quad G(s,0) = s^{n_0}.$$

On peut la résoudre et trouver :

$$G(s,t) = \Big(\frac{(\lambda s - \mu)e^{-t(\lambda-\mu)} - \mu(s-1)}{(\lambda s - \mu)e^{-t(\lambda-\mu)} - \lambda(s-1)}\Big)^{n_0}.$$

Chapitre 5

Chaînes de Markov

Les *chaînes de Markov* constituent l'exemple le plus simple des processus stochastiques, lorsque dans l'étude d'une suite de variables aléatoires, on abandonne l'hypothèse d'indépendance. Il s'agit d'un processus à *temps discret* — d'où le nom de « chaîne ». Il y a les chaînes de Markov dont l'ensemble des états est *continu* — leur étude n'est pas abordée ici — et les chaînes dont l'ensemble des états est *discret*. Celles-ci font l'objet du présent chapitre et du suivant. L'ensemble E des états est donc soit $\mathbb{N}$, soit une partie finie de celui-ci.

5.1 DÉFINITION ET PREMIÈRES PROPRIÉTÉS

Soit (X_n) $(n \geq 0)$ une suite de variables aléatoires à valeurs dans l'ensemble E des états, supposé égal à $\mathbb{N}$. On dit que cette suite est une *chaîne de Markov*, si pour tout $n \geq 1$ et toute suite $(i_0, \ldots, i_{n-1}, i, j)$ d'éléments de E, pour laquelle la probabilité $\mathrm{P}\{X_0 = i_0, \ldots, X_{n-1} = i_{n-1}, X_n = i\}$ est strictement positive, on a la relation suivante entre probabilités conditionnelles :

$$\mathrm{P}\{X_{n+1} = j \mid X_0 = i_0, \ldots, X_{n-1} = i_{n-1}, X_n = i\} = \mathrm{P}\{X_{n+1} = j \mid X_n = i\}. \quad (5.1.1)$$

Autrement dit, dans l'évolution au cours du temps, l'état du processus à l'instant $(n+1)$ ne dépend que de celui à l'instant n précédent, mais non de ses états antérieurs. Le processus est *sans mémoire* ou *non héréditaire*.

Définition 5.1.1. La chaîne de Markov est dite *homogène* (dans le temps), si la probabilité précédente *ne dépend pas de n*. Soit

$$p_{i,j} := \mathrm{P}\{X_{n+1} = j \mid X_n = i\} \quad (n \geq 0) \tag{5.1.2}$$

cette probabilité ; on l'appelle la *probabilité de passage de l'état i à l'état j*, en une étape, ou en une opération, ou encore, en une transition.

Définition 5.1.2. La matrice

$$\mathcal{P} = \begin{pmatrix} p_{0,0} & p_{0,1} & p_{0,2} & \cdots \\ p_{1,0} & p_{1,1} & p_{1,2} & \cdots \\ \vdots & \vdots & \vdots & \vdots\vdots\vdots \end{pmatrix}$$

dont les coefficients sont les probabilités de transition $p_{i,j}$ est appelée matrice *de passage* (ou *de transition*) de la chaîne. C'est une matrice finie ou dénombrable, suivant que l'ensemble des états est fini ou dénombrable.

Propriété 5.1.1. *Toute matrice de transition $\mathcal{P} = (p_{i,j})$ $((i,j) \in E^2)$ vérifie les propriétés suivantes :*

(1) *pour tout couple (i,j), on a : $p_{ij} \geq 0$;*

(2) *pour tout $i \in E$, on a $\sum_{j \in E} p_{ij} = 1$.*

Démonstration. Les nombres $p_{i,j}$ sont des probabilités, donc des nombres positifs. Par ailleurs, pour chaque $i \in E$, l'application $B \mapsto \sum_{j \in B} p_{ij}$ définit une mesure de probabilité sur E. □

Remarque. Une matrice $\mathcal{P}$, qui vérifie les conditions (1) et (2) de la propriété précédente, est appelée *matrice stochastique*. stochastique (matrice)

Propriété 5.1.2. *Soit $\mathcal{P}$ une matrice de transition. Alors*

(1) *$\mathcal{P}$ admet la valeur propre* 1 *;*

(2) *on peut associer à cette valeur propre le vecteur propre V ayant toutes ses composantes égales à* 1.

Démonstration. En effet, en considérant V comme un vecteur-colonne, on a : $\mathcal{P} \cdot V = V$ si et seulement si, pour tout $i \in E$, la relation suivante est satisfaite : $\sum_{j \in E} p_{i,j} v_j = v_i$. Il suffit donc, pour tout $i \in E$, de prendre $v_i := 1$. □

Graphe associé à une matrice de transition. À toute matrice de transition, on peut associer son *graphe*. Les *sommets* du graphe sont les différents *états* de la chaîne. Il y a une *flèche*, étiquetée $p_{i,j}$, entre le sommet étiqueté i et le sommet étiqueté j si et seulement si la probabilité de transition de l'état i à l'état j est *strictement* positive : $p_{i,j} > 0$.

Lorsque l'ensemble des états est fini, cette présentation de la matrice de transition par son graphe est particulièrement utile et parlante.

5.2 EXEMPLES DE CHAÎNES DE MARKOV

Il y a des modèles classiques de chaînes de Markov homogènes, souvent rencontrés, avec lesquels il est bon de se familiariser dès le début. En voici quelques-uns.

5.2.1 La chaîne à deux états

En excluant le cas trivial de la matrice-unité, la matrice de transition correspondante est de la forme

$$\mathcal{P} = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix} \qquad (0 < \alpha, \beta \leq 1).$$

Les calculs sont explicites. Pour tout $n \geq 0$, on peut évaluer la $n^{\text{ième}}$ puissance $\mathcal{P}^n$, ainsi que la valeur limite $\lim_n \mathcal{P}^n$ (*cf.* Exercices 1 et 2).

Le graphe associé est très simple :

$$1-\alpha \circlearrowleft \overset{1}{\bullet} \underset{\beta}{\overset{\alpha}{\rightleftarrows}} \overset{2}{\bullet} \circlearrowright 1-\beta$$

5.2.2 Le modèle de diffusion d'Ehrenfest

Deux urnes A et B contiennent, à elles deux, a boules, numérotées de 1 à a. À chaque instant, on choisit un nombre de 1 à a, avec une probabilité $1/a$. Si ce nombre est i, on change d'urne la boule numérotée i.

L'ensemble des états est l'ensemble $E = \{0, 1, \ldots, a\}$. Le processus est dit *être dans l'état* j si l'urne A contient j boules. Dans ces conditions, si le processus est dans l'état 0 (l'urne A est vide) (resp. l'état a, à savoir l'urne B est vide), la probabilité est égale à 1 qu'il passe dans l'état 1 (resp. l'état $(a-1)$). La *matrice de transtion* est donc donnée par

$$\mathcal{P} = \begin{pmatrix} 0 & 1 & 0 & 0 & \ldots & 0 & 0 & 0 \\ 1/a & 0 & (a-1)/a & 0 & \ldots & 0 & 0 & 0 \\ 0 & 2/a & 0 & (a-2)/a & \ldots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \ldots & (a-1)/a & 0 & 1/a \\ 0 & 0 & 0 & 0 & \ldots & 0 & 1 & 0 \end{pmatrix}$$

et le *graphe* associé par :

$$\underset{0}{\bullet} \underset{1/a}{\overset{1}{\rightleftarrows}} \underset{1}{\bullet} \underset{2/a}{\overset{(a-1)/a}{\rightleftarrows}} \underset{2}{\bullet} \underset{3/a}{\overset{(a-2)/a}{\rightleftarrows}} \cdots \underset{(a-2)/a}{\overset{3/a}{\rightleftarrows}} \overset{a-2}{\bullet} \underset{(a-1)/a}{\overset{2/a}{\rightleftarrows}} \overset{a-1}{\bullet} \underset{1}{\overset{1/a}{\rightleftarrows}} \overset{a}{\bullet}$$

Pour $n \geq 0$, désignons par X_n le nombre de boules dans l'urne A à l'instant n. Si $X_0 = a$, alors le processus (X_n) $(n \geq 0)$ décrit la « diffusion » d'un gaz de A vers B.

5.2.3 Promenade au hasard sur $\mathbb{Z}$ (Lord Rayleigh)

Considérons une chaîne de Markov, homogène dans le temps, caractérisée par l'ensemble des états $E := \mathbb{Z} = \{0, \pm 1, \pm 2, \dots\}$ et la matrice de transition $\mathcal{P} = (p_{i,j})$ $(i, j \in \mathbb{Z})$, avec, pour tout $(i, j) \in \mathbb{Z}^2$,

$$p_{i,j} = \begin{cases} p, & \text{si } j = i + 1\,; \\ 1 - p, & \text{si } j = i - 1\,; \\ 0, & \text{dans les autres cas}\,; \end{cases}$$

où p est un nombre fixé tel que $0 < p < 1$. Un tel processus est appelé *promenade aléatoire sur* $\mathbb{Z}$. Son graphe peut être décrit comme suit :

$$\cdots \bullet_{i-1} \underset{1-p}{\overset{p}{\rightleftarrows}} \bullet_{i} \underset{1-p}{\overset{p}{\rightleftarrows}} \bullet_{i+1} \cdots$$

5.2.4 Le modèle de la ruine du joueur

Un joueur A joue contre un joueur B une suite de parties de « pile » ou « face », *indépendantes.* La somme totale de leurs fortunes est de a euros. À chaque partie, on convient que le joueur A gagne un euro (que lui donne le joueur B) avec une probabilité p et perd un euro (qu'il rend donc à B) avec une probabilité q $(0 < p < 1;\ q = 1 - p)$. Le jeu s'arrête dès que l'un des joueurs est ruiné.

Pour chaque $n \geq 0$, on désigne par X_n la fortune du joueur A à l'issue de la $n^{\text{ième}}$ partie. La suite (X_n) $(n \geq 0)$ est une chaîne de Markov, dont l'ensemble des états est $E := \{0, 1, \dots, a\}$. Sa matrice de transition est donnée par

$$\mathcal{P} = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 & 0 & 0 \\ q & 0 & p & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & q & 0 & p \\ 0 & 0 & 0 & \dots & 0 & 0 & 1 \end{pmatrix}$$

et son graphe par :

$$1 \circlearrowleft \bullet_0 \overset{}{\underset{q}{\leftarrow}} \bullet_1 \underset{q}{\overset{p}{\rightleftarrows}} \bullet_2 \underset{q}{\overset{p}{\rightleftarrows}} \cdots \underset{q}{\overset{p}{\rightleftarrows}} \bullet_{a-2} \underset{q}{\overset{p}{\rightleftarrows}} \bullet_{a-1} \overset{p}{\rightarrow} \bullet_a \circlearrowright 1$$

Les états 0 (ruine de A) et a (ruine de B) sont dits *absorbants.*

5.2.5 Le modèle de bonus-malus

Une compagnie d'assurances classe les niveaux de *bonus-malus* de ses clients suivant les entiers naturels : $0, 1, 2, \dots$ Le niveau 0 est le plus avantageux pour l'assuré. Soit $0 \leq i \leq j$. Si le niveau de *bonus-malus* de celui-ci à l'instant n est i, son niveau à l'instant suivant $(n+1)$ est j si entre les deux instants il a eu $(j - i)$ accidents. Il ne rétrogradera jamais ! On désigne par (X_n) $(n \geq 0)$

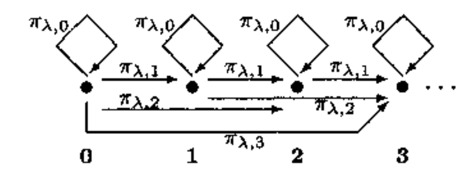
la suite des niveaux de *bonus-malus* qu'un assuré donné parcourt au cours du temps.

Convenons, d'autre part, que l'unité de temps est l'année et que le nombre d'accidents provoqués par l'assuré en une année suit la loi de Poisson de paramètre λ. Alors la probabilité que l'assuré passe du niveau i au niveau j avec $j \ge i$ est égale à $\pi_{\lambda,j-i} := e^{-\lambda}\lambda^{j-i}/(j-i)!$

D'après ces hypothèses, la suite (X_n) $(n \ge 0)$ est une chaîne de Markov, dont l'ensemble des états est $E = \mathbb{N}$, la matrice de transition est triangulaire supérieure, donnée par

$$\mathcal{P} = \begin{pmatrix} \pi_{\lambda,0} & \pi_{\lambda,1} & \pi_{\lambda,2} & \cdots \\ 0 & \pi_{\lambda,0} & \pi_{\lambda,1} & \cdots \\ 0 & 0 & \pi_{\lambda,0} & \cdots \\ \vdots & \vdots & \vdots & \vdots\vdots\vdots \end{pmatrix}.$$

Son graphe est le suivant :



5.2.6 Évolution déterministe

À toute application f d'un ensemble E fini ou dénombrable dans lui-même, on peut faire correspondre une matrice de transition $\mathcal{P}_f = (p_{i,j})$ $((i,j) \in E^2)$ définie par :

$$p_{i,j} := \begin{cases} 1, & \text{si } f(i) = j\,; \\ 0, & \text{autrement.} \end{cases}$$

Si f, g sont deux applications de E dans E, la matrice de transition $\mathcal{P}_{f\circ g}$ associée à $f \circ g$ n'est autre que le produit matriciel $\mathcal{P}_f\,\mathcal{P}_g$. En particulier, $\mathcal{P}_{f^n} = (\mathcal{P}_f)^n$.

La suite des ensembles E, $f(E)$, $f^2(E)$, ... est *décroissante* pour la relation d'inclusion. Lorsque E est *fini*, il existe $E_0 \subset E$ tel que $f^n(E) = E_0$ à partir d'un certain rang et la restriction de f à l'ensemble E_0 est une *permutation* de E_0. Comme on peut décomposer cette permutation en *cycles* disjoints $E_0 = C_1 + \cdots + C_r$, l'ensemble E lui-même, se partitionne en *orbites* disjointes $E = O_1 + \cdots + O_r$, avec $C_i \subset O_i$ pour $i = 1, \ldots, r$. À partir d'un certain rang n, l'élément $f^n(e)$ parcourt donc l'un des cycles C_i.

Si (X_n) $(n \ge 0)$ est une suite de variables aléatoires, ayant E pour ensemble d'états et $\mathcal{P}_f$ pour matrice de transition. Le comportement purement *déterministe* de la chaîne est complètement décrit par la suite $(\mathcal{P}_{f^n})$ $(n \ge 0)$. En particulier, asymptotiquement, la chaîne a un comportement périodique, elle reste dans l'un des cycles C_i. Cette évolution déterministe donne une

idée des phénomènes de périodicité que l'on rencontre dans les processus de Markov plus généraux.

Le graphe associé à $\mathcal{P}_f$ n'est autre que le graphe usuel associé à l'application f (*cf.* Exercice 5.4). Notons que si E est fini et si f est une permutation de E, la matrice $\mathcal{P}_f$ est une matrice dite *de permutation*, qui n'a que des 0 et des 1, avec un et un seul 1 dans chaque ligne et dans chaque colonne.

5.3 LA RELATION DE CHAPMAN-KOLMOGOROV

On a coutume d'appeler ainsi la propriété qui relie, pour une chaîne de Markov homogène, les probabilités de transition en n étapes aux probabilités en une seule étape. On conserve les notations des paragraphes précédents : (X_n) $(n \geq 0)$ est une chaîne de Markov homogène, dont l'ensemble des états est E et la matrice de transition $\mathcal{P} = (p_{i,j})$ $((i,j) \in E^2)$. Pour $n \geq 0$ et $i, j \in E$, on désigne par $p_{i,j}^{(n)}$ la probabilité, partant de l'état i en l'instant 0, d'être dans l'état j en l'instant n ; en d'autres termes, on pose :

$$p_{i,j}^{(n)} := \mathrm{P}\{X_n = j \mid X_0 = i\}. \tag{5.3.1}$$

On pose également : $\mathcal{P}^{(n)} := (p_{i,j}^{(n)})$ $((i,j) \in E^2)$.

Théorème 5.3.1 (Relation de Chapman-Kolmogorov). *Pour tout $n \geq 0$, la matrice de transition en n étapes est égale à la puissance $n^{\text{ième}}$ de la matrice de transition en une étape :*

$$\mathcal{P}^{(n)} = (\mathcal{P})^n.$$

Démonstration. Le résultat est vrai pour $n = 0$ puisque $\mathcal{P}^0 = I$ (la matrice-identité) et pour $n = 1$, puisque $\mathcal{P}^{(1)} = \mathcal{P}$. Supposons $n \geq 2$; par récurrence, $\mathcal{P}^n = \mathcal{P}^{n-1}\mathcal{P} = \mathcal{P}^{(n-1)}\mathcal{P}^{(1)}$. Par conséquent, si l'on désigne par $p_{i,j}^n$ le coefficient en (i,j) de la matrice $\mathcal{P}^n$, on a :

$$\begin{aligned} p_{i,j}^n &= \sum_{k \in E} p_{i,k}^{(n-1)} p_{k,j} \\ &= \sum_{k \in E} \mathrm{P}\{X_{n-1} = k \mid X_0 = i\}\, \mathrm{P}\{X_1 = j \mid X_0 = k\} \\ &= \sum_{k \in E} \mathrm{P}\{X_{n-1} = k \mid X_0 = i\}\, \mathrm{P}\{X_n = j \mid X_{n-1} = k\} \end{aligned}$$

[car la chaîne est homogène.]

Considérons l'évènement $A(i_1, \ldots, i_{n-2}) := \{X_1 = i_1, \ldots, X_{n-2} = i_{n-2}\}$. Alors

$$\mathrm{P}\{X_n = j \mid X_{n-1} = k\} = \mathrm{P}\{X_n = j \mid X_{n-1} = k, A(i_1, \ldots, i_{n-2}), X_0 = i\},$$

car la chaîne est de Markov et

$$\mathrm{P}\{X_{n-1} = k \mid X_0 = i\} = \sum_{i_1, \ldots, i_{n-2}} \mathrm{P}\{X_{n-1} = k, A(i_1, \ldots, i_{n-2}) \mid X_0 = i\},$$

puisque les évènements $A(i_1, \dots, i_{n-2})$ forment un système complet d'évènements. On en tire :

$$\begin{aligned}
p_{i,j}^n &= \sum_{k\in E} \sum_{i_1,\dots,i_{n-2}} \mathrm{P}\{X_{n-1}=k, A(i_1,\dots,i_{n-2}) \,|\, X_0=i\} \\
&\qquad\qquad \times \mathrm{P}\{X_n=j \,|\, X_{n-1}=k\} \\
&= \sum_{k\in E} \sum_{i_1,\dots,i_{n-2}} \mathrm{P}\{X_{n-1}=k, A(i_1,\dots,i_{n-2}) \,|\, X_0=i\} \\
&\qquad\qquad \times \mathrm{P}\{X_n=j \,|\, X_{n-1}=k, A(i_1,\dots,i_{n-2}), X_0=i\} \\
&= \sum_{k\in E} \sum_{i_1,\dots,i_{n-2}} \mathrm{P}\{X_n=j, X_{n-1}=k, A(i_1,\dots,i_{n-2}) \,|\, X_0=i\} \\
&= \mathrm{P}\{X_n=j \,|\, X_0=i\} = p_{i,j}^{(n)}. \quad \square
\end{aligned}$$

Corollaire 5.3.2. *Pour tout $n \geq 0$, la matrice $\mathcal{P}^{(n)}$ est stochastique.*

Démonstration. En effet, pour tout $i \in E$, on a :

$$\sum_{j\in E} p_{i,j}^{(n)} = \sum_{j\in E} \mathrm{P}\{X_n=j \,|\, X_0=i\} = 1. \quad \square$$

Ainsi, pour tout $n \geq 1$ fixé, la suite (X_{nm}) $(m \geq 0)$ est une chaîne de Markov, ayant même ensemble d'états E et dont la matrice de transition est $\mathcal{P}^n$.

Corollaire 5.3.3. *Pour tout $(i,j) \in E^2$ et tout couple (m,n) d'entiers positifs, on a l'identité :*

$$\mathrm{P}\{X_{m+n}=j \,|\, X_0=i\} = \sum_{k\in E} \mathrm{P}\{X_m=k \,|\, X_0=i\}\, \mathrm{P}\{X_n=j \,|\, X_0=k\},$$

ou encore

$$p_{i,j}^{(m+n)} = \sum_{k\in E} p_{i,k}^{(m)}\, p_{k,j}^{(n)}. \tag{5.3.2}$$

Démonstration. Cette identité résulte immédiatement de l'associativité du produit matriciel : $\mathcal{P}^{(m+n)} = \mathcal{P}^{m+n} = \mathcal{P}^m\mathcal{P}^n = \mathcal{P}^{(m)}\mathcal{P}^{(n)}$. $\square$

Proposition 5.3.4. *Soient $n \geq 0$, $r \geq 1$ deux entiers. Alors*

$$\begin{aligned}
\mathrm{P}\{X_{n+1}=j_{n+1}, \dots, X_{n+r}=j_{n+r} \,|\, X_0=i_0, \dots, X_n=i_n\} \\
= p_{i_n,j_{n+1}}\, p_{j_{n+1},j_{n+2}} \cdots p_{j_{n+r-1},j_{n+r}}.
\end{aligned} \tag{5.3.3}$$

Démonstration. Lorsque $r=1$, l'identité (5.3.3) se réduit à (5.1.1). Il suffit donc de procéder par récurrence sur r. Pour $r \geq 2$, posons :

$$\begin{aligned}
A_1 &:= \{X_0=i_0, \dots, X_n=i_n\}; \\
A_2 &:= \{X_{n+1}=j_{n+1}, \dots, X_{n+r-1}=j_{n+r-1}\}; \\
A_3 &:= \{X_{n+r}=j_{n+r}\}.
\end{aligned}$$

D'après (5.1.1), on a : $\mathrm{P}(A_3 \,|\, A_1A_2) = p_{j_{n+r-1},j_{n+r}}$; de plus, $P(A_2 \,|\, A_1) = p_{i_n,j_{n+1}}\, p_{j_{n+1},j_{n+2}} \cdots p_{j_{n+r-2},j_{n+r-1}}$, par récurrence sur r. On conclut alors, en utilisant l'identité : $\mathrm{P}(A_2A_3 \,|\, A_1) = \mathrm{P}(A_3 \,|\, A_1A_2)\, \mathrm{P}(A_2 \,|\, A_1)$. $\square$

5.4 MESURE DE PROBABILITÉ D'UNE CHAÎNE DE MARKOV

Soit (X_n) $(n \geq 0)$ une chaîne de Markov, homogène, admettant l'ensemble E, fini ou dénombrable, comme ensemble des états et la matrice $\mathcal{P} = (p_{i,j})$ $((i,j) \in E^2)$ comme matrice de transition. Supposons, de plus, donnée une mesure de probabilité μ sur E, autrement dit, une suite (μ_i) $(i \in E)$ de nombres positifs tels que $\sum_{i \in E} \mu_i = 1$. Le but de ce paragraphe est de montrer que la matrice de transition $\mathcal{P}$ et la mesure de probabilité μ permettent de définir une *mesure de probabilité* P sur l'espace probabilisable $(\Omega, \mathfrak{A})$ sur lequel sont définies les variables X_n.

5.4.1 Mesure de probabilité et loi initiale

Définition 5.4.1. On dit que P est la *mesure de probabilité* de cette chaîne, si les propriétés suivantes sont vérifiées :

(i) pour tout état i, on a : $\mathrm{P}\{X_0 = i\} = \mu_i$; on dit alors que μ est la *loi initiale* de la chaîne ;

(ii) pour tout couple (i,j) d'états, on a : $\mathrm{P}\{X_1 = j \mid X_0 = i\} = p_{i,j}$.

Dans ces conditions, l'évaluation suivante résulte de la Proposition 5.3.4, formule (5.3.3), pour $n = 0$ et $r \geq 1$:

$$\begin{aligned} \mathrm{P}\{X_0 = i_0, X_1 = j_1, \dots, X_r = j_r\} \\ &= \mathrm{P}\{X_0 = i_0\}\,\mathrm{P}\{X_1 = j_1, \dots, X_r = j_r \mid X_0 = i_0\} \\ &= \mu_{i_0}\, p_{i_0,j_1} \cdots p_{j_{r-1},j_r}. \end{aligned} \tag{5.4.1}$$

On note aussi P^μ cette mesure de probabilité de la chaîne, lorsque l'on veut faire ressortir la dépendance de P de la loi initiale μ.

5.4.2 Le processus canonique

Considérons l'ensemble $E^{\mathbb{N}}$ des suites infinies dont les termes sont pris dans un ensemble dénombrable E. Pour $n = 0, 1, \dots$, la $n^{\text{ième}}$ *projection* de $E^{\mathbb{N}}$ sur E est définie comme l'application X_n qui envoie l'élément $\omega = (i_0, i_1, \dots, i_n, \dots)$ sur $X_n(\omega) := i_n$.

Pour décrire une chaîne de Markov qui possède la mesure de probabilité P^μ, on prend pour Ω l'ensemble $E^{\mathbb{N}}$, pour suite (X_n) $(n \geq 0)$ la suite des projections, pour tribu $\mathfrak{A}$ la tribu engendrée par tous les X_n (*cf.* § 1.12), c'est-à-dire la tribu engendrée par les évènements de la forme $\{X_n = i\}$. On peut montrer alors qu'il existe une et une seule loi de probabilité $\mathrm{P} := \mathrm{P}^\mu$ sur $(\Omega, \mathfrak{A})$ telle que les relations (5.4.1) soient satisfaites. (Voir [FF1], chap. 10, Exercices 1–6, pour une démonstration dans le cas où E est fini.) Le processus ainsi construit sur $E^{\mathbf{N}}$ est dit *canonique*.

5.4.3 Formules de conditionnement

Dans les énoncés suivants, A désigne un évènement appartenant à la tribu $\mathfrak{A}_n$ $(n \geq 0)$, une tribu qui est engendrée par le vecteur $(X_0, \dots, X_n)$. Comme

vu dans la Proposition 1.12.2, l'évènement A est une réunion dénombrable d'évènements, disjoints deux à deux, de la forme $\{X_0 = i_0, \dots, X_n = i_n\}$. L'évènement $\{A, X_n = i\}$ est alors un évènement de la tribu $\mathfrak{A}_n$. S'il n'est pas vide, on a forcément $A \subset \{X_n = i\}$; le « présent » $\{X_n = i\}$ est donc fixé.

Proposition 5.4.1. *Soient $n \geq 0$ et A un évènement appartenant à la tribu $\mathfrak{A}_n$. Si $\mathrm{P}\{A, X_n = i\} > 0$, on a :*

$$\mathrm{P}\{X_{n+1} = j \mid A, X_n = i\} = \mathrm{P}\{X_{n+1} = j \mid X_n = i\} = p_{i,j}. \tag{5.4.2}$$

Démonstration. Pour $n = 0$, la probabilité conditionnelle dans (5.4.2) n'est définie que pour $A = \{X_0 = i\}$ et dans ce cas elle vaut $p_{i,j}$, si $\mathrm{P}\{X_0 = i\} > 0$. On dit qu'un évènement A, appartenant à la tribu $\mathfrak{A}_n$, a la propriété $(\mathcal{H})$ si $\mathrm{P}\{A, X_n = i\} > 0$ et si la relation (5.4.2) est satisfaite. Considérons une suite d'événèments (A_k), *disjoints deux à deux*, ayant la propriété $(\mathcal{H})$. Alors

$$\begin{aligned}\mathrm{P}\{X_{n+1} = j, \sum_k A_k, X_n = i\} &= \sum_k \mathrm{P}\{X_{n+1} = j, A_k, X_n = i\} \\ &= \sum_k \mathrm{P}\{X_{n+1} = j \mid A_k, X_n = i\}\mathrm{P}\{A_k, X_n = i\} \\ &= p_{i,j} \sum_k \mathrm{P}\{A_k, X_n = i\} = p_{i,j}\mathrm{P}\{\sum_k A_k, X_n = i\}.\end{aligned}$$

Comme la probabilité $\mathrm{P}\{\sum_k A_k, X_n = i\}$ est strictement positive, la réunion $\sum_k A_k$ a encore la propriété $(\mathcal{H})$. D'après la définition-même (5.1.1) de la propriété de Markov, chaque évènement de la forme $\{X_0 = i_0, \dots, X_{n-1} = i_{n-1}, X_n = i_n\}$, satisfaisant $\mathrm{P}\{X_0 = i_0, \dots, X_{n-1} = i_{n-1}, X_n = i_n, X_n = i\} > 0$ (ce qui implique $i_n = i$), a la propriété $(\mathcal{H})$. Comme tout évènement A de $\mathfrak{A}_n$, tel que $\mathrm{P}\{A, X_n = i\} > 0$, est réunion dénombrable de tels évènements, il satisfait aussi la relation (5.4.2). □

Remarque. Pour que la propriété (5.1.1) puisse se prolonger à un évènement de la tribu $\mathfrak{A}_n$, comme dans (5.4.2), il faut que le *présent* soit de la forme $\{X_n = i\}$, où i est un *élément* de l'ensemble des états. La formule ne serait plus valable, si ce présent était de la forme $\{X_n \in B\}$ avec B non-singleton.

Considérons, en effet, la promenade au hasard sur $\mathbb{Z}$ (*cf.* § 5.2.3) avec $\frac{1}{2}(\varepsilon_{-2} + \varepsilon_2)$ comme loi initiale. Alors $\mathrm{P}\{X_0 = 2, X_1 = \pm 1, X_2 = 0\} = \frac{1}{2}q^2$, puis $\mathrm{P}\{X_0 = 2, X_1 = \pm 1\} = \frac{1}{2}q$ et $\mathrm{P}\{X_1 = \pm 1\} = \frac{1}{2}(p + q) = \frac{1}{2}$; enfin $\mathrm{P}\{X_1 = \pm 1, X_2 = 0\} = \frac{1}{2}(p^2 + q^2)$. On en déduit : $\mathrm{P}\{X_2 = 0 \mid X_0 = 2, X_1 \in \{-1, +1\}\} = q$, alors que $\mathrm{P}\{X_2 = 0 \mid X_1 \in \{-1, +1\}\} = p^2 + q^2$. Ces deux probabilités sont donc différentes si $q \neq 1, \frac{1}{2}$.

On peut aussi prolonger la formule (5.4.2), en considérant non plus des évènements futurs de la forme $\{X_{n+1} = j\}$, mais des *suites finies d'évènements ponctuels*, tous postérieurs à $(n + 1)$ et obtenir ainsi une extension des formules (5.3.3) et (5.4.2).

Proposition 5.4.2. *Soient $n \geq 0$, $r \geq 1$ deux entiers et A un évènement appartenant à $\mathfrak{A}_n$. Si $\mathrm{P}\{A, X_n = i\} > 0$, on a :*

$$\mathrm{P}\{X_{n+1} = j_{n+1}, \ldots, X_{n+r} = j_{n+r} \,|\, A, X_n = i_n\} = p_{i_n,j_{n+1}}\, p_{j_{n+1},j_{n+2}} \cdots p_{j_{n+r-1},j_{n+r}}. \quad (5.4.3)$$

Démonstration. La démonstration est tout à fait analogue à celle de la Proposition 5.4.1, en vertu de la Proposition 5.3.4. □

Lorsque $r = 1$, la formule suivante est identique à (5.4.2) ou à (5.4.3). Pour $r \geq 2$, elle se déduit de cette dernière formule par sommation par rapport à $j_{n+1}, \cdots, j_{n+r-1}$:

$$\mathrm{P}\{X_{n+r} = j \,|\, A, X_n = i\} = p_{i,j}^{(r)}. \quad (5.4.4)$$

En particulier, pour $A = \Omega$, on obtient :

$$\mathrm{P}\{X_{n+r} = j \,|\, X_n = i\} = p_{i,j}^{(r)}, \quad (5.4.5)$$

ce qui montre que l'homogénéité supposée pour la transition en une étape est encore vraie pour les transitions en r étapes.

Comme le membre de droite de (5.4.3) ne dépend pas de A, on obtient :

$$\begin{aligned} \mathrm{P}\{X_{n+1} = j_{n+1}, \ldots, X_{n+r} = j_{n+r} \,|\, A, X_n = i_n\} &= \mathrm{P}\{X_{n+1} = j_{n+1}, \ldots, X_{n+r} = j_{n+r} \,|\, X_n = i_n\} \\ &= p_{i_n,j_{n+1}}\, p_{j_{n+1},j_{n+2}} \cdots p_{j_{n+r-1},j_{n+r}}. \quad (5.4.6) \end{aligned}$$

Pour démontrer, enfin, que la première de ces identités se prolonge à tous les éléments de la tribu $\mathfrak{A}_{n+1,\infty}$, on procède comme suit. D'abord, la dite identité se prolonge à tous les éléments de la tribu $\mathfrak{A}_{n+1,n+r}$ engendrée par le vecteur $(X_{n+1}, \ldots, X_{n+r})$, puisque tous les éléments de cette tribu sont des réunions dénombrables d'évènements de la forme $\{X_{n+1} = j_{n+1}, \ldots, X_{n+r} = j_{n+r}\}$ (*cf.* Proposition 1.12.1). Autrement dit, pour tout $B \in \mathfrak{A}_{n+1,n+r}$, on a :

$$\mathrm{P}\{B \,|\, A, X_n = i_n\} = \mathrm{P}\{B \,|\, X_n = i_n\}. \quad (5.4.7)$$

L'identité (5.4.7) est encore valable pour tous les éléments B de l'*algèbre* $\mathcal{A}_{n+1,\infty}$, définie comme la réunion de toutes les tribus $\mathfrak{A}_{n+1,n+r}$ $(r \geq 1)$. Pour démontrer qu'elle est finalement valable pour les éléments de la *tribu* $\mathfrak{A}_{n+1,\infty}$, on utilise l'argument suivant. D'abord, cette tribu est aussi engendrée par l'algèbre $\mathcal{A}_{n+1,\infty}$. Ensuite, on vérifie que si l'identité (5.4.7) est vraie pour les éléments d'une suite *monotone* d'évènements, elle est aussi vraie pour la limite de cette suite, un résultat immédiat à vérifier. L'identité (5.4.7) est donc valable pour les éléments d'une *classe monotone*, qui contient l'algèbre $\mathcal{A}_{n,\infty}$. Elle est donc aussi vraie pour les éléments de la tribu engendrée $\mathfrak{A}_{n+1,\infty}$ (*cf.* [FF1], chap. 2, Proposition 4.2).

5.5 CLASSIFICATION DES ÉTATS; DÉCOMPOSITION EN CLASSES

Les états d'une chaîne de Markov se répartissent en *classes* que l'on définit à partir de la matrice de transition.

Définition 5.5.1. On dit que l'état j est *accessible à partir de* l'état i, ou est *conséquent de* l'état i, s'il existe un entier $n \geq 0$ tel que $p_{i,j}^{(n)} > 0$. On écrit : $i \rightsquigarrow j$.

Propriété 5.5.1. *La relation d'accessibilité entre états est réflexive et transitive.*

Démonstration. Comme $p_{i,i}^{(0)} = \mathrm{P}\{X_0 = i \,|\, X_0 = i\} = 1$ pour tout état i, on a bien $i \rightsquigarrow i$. Ensuite, si $i \rightsquigarrow l$ et $l \rightsquigarrow j$, alors $p_{i,l}^{(m)} > 0$ et $p_{l,j}^{(n)} > 0$ pour certains entiers $m, n \geq 0$. D'après la relation (5.3.2), on en tire : $p_{i,j}^{(m+n)} = \sum_{k \in E} p_{i,k}^{(m)} p_{k,j}^{(n)} \geq p_{i,l}^{(m)} p_{l,j}^{(n)} > 0$, d'où $i \rightsquigarrow j$. □

Propriété 5.5.2. *Soient i, j deux états; les deux propriétés suivantes sont équivalentes :*

(a) *l'état j est accessible à partir de l'état i, soit $i \rightsquigarrow j$;*

(b) *le processus, partant de i, passe par j avec une probabilité strictement positive.*

Démonstration. D'abord, (a) $\Rightarrow$ (b) est évident. Montrons que $\overline{(\mathrm{a})} \Rightarrow \overline{(\mathrm{b})}$. Sous l'hypothèse $\overline{(\mathrm{a})}$, pour tout $n \geq 0$, on a : $p_{i,j}^{(n)} = 0$. Soit A l'évènement « le processus passe par j ». Alors $\mathrm{P}\{A \,|\, X_0 = i\} = \mathrm{P}\{\sum_{n \geq 0} \{X_n = j\} \,|\, X_0 = i\} \leq \sum_{n \geq 0} \mathrm{P}\{X_n = j \,|\, X_0 = i\} = \sum_{n \geq 0} p_{i,j}^{(n)} = 0$, d'où la propriété $\overline{(\mathrm{b})}$. □

Définition 5.5.2. On dit que deux états i et j *communiquent* et l'on écrit $i \leftrightsquigarrow j$, si on a à la fois : $i \rightsquigarrow j$ et $j \rightsquigarrow i$.

Propriété 5.5.3. *La relation de communication entre états est une relation d'équivalence.*

Démonstration. Les propriétés de réflexivité et de transitivité déjà vérifiées pour la relation d'accessibilité restent naturellement encore valables pour la relation de communication. Enfin, cette dernière relation est symétrique par définition-même. □

Remarque. En raison du fait que, pour tout i, on a $p_{i,i}^{(0)} = 1$, tout état communique avec lui-même. Un état est appelé *état de retour*, s'il existe $n \geq 1$ tel que $p_{i,i}^{(n)} > 0$. Il existe des états i tels que pour tout $n \geq 1$ (donc 0 exclu) on ait $p_{i,i}^{(n)} = 0$. De tels états sont appelés *états de non-retour*. Par exemple,

l'état 1 dans la chaîne de Markov, telle que $E = \{0,1\}$ et $\mathcal{P} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}$ est un état de non-retour.

Pour la relation de communication, l'ensemble E des états se *partitionne* en *classes* d'équivalence, disjointes et non vides, dites *classes indécomposables.* Si C_1 et C_2 sont deux classes distinctes, on peut éventuellement aller, disons, de C_1 à C_2, mais on ne peut alors retourner de C_2 à C_1. En revanche, tous les états d'une même classe communiquent.

Certaines classes peuvent ne comporter qu'un seul élément ; ce sont les *singletons.* Comme exemples, mentionnons :

un *état de non-retour* i : $p_{i,i}^{(0)} = 1$, $p_{i,i}^{(n)} = 0$ pour $n \geq 1$;

un *état absorbant* i : $p_{i,i}^{(0)} = 1$, $p_{i,i}^{(n)} = 1$ pour $n \geq 1$.

Dans l'exemple juste mentionné, l'état 0 est absorbant.

Définition 5.5.3. S'il n'y a qu'une seule classe pour la relation de communication, autrement dit, si tous les états communiquent entre eux, la chaîne est dite *irréductible.*

Exemple 1. Considérons la chaîne de Markov, dont l'ensemble des états est $E = \{0,1,2\}$, la matrice de transition

$$\mathcal{P} = \begin{pmatrix} 1/2 & 1/2 & 0 \\ 1/2 & 1/4 & 1/4 \\ 0 & 1/3 & 2/3 \end{pmatrix}$$ et le graphe

Cette chaîne est *irréductible.* Tous les états communiquent. Bien que $p_{0,2} = p_{2,0} = 0$ et qu'il n'y a donc pas de flèche entre les états 0 et 2, on a, en revanche, pour tout $n \geq 2$ et tout couple d'états (i,j), l'inégalité *stricte* : $p_{i,j}^{(n)} > 0$.

Exemple 2. Considérons la chaîne de Markov, dont l'ensemble des états est $E = \{0,1,2,3\}$, la matrice de transition

$$\mathcal{P} = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$ et le graphe

La chaîne comporte trois classes $\{0,1\}$, $\{2\}$ et $\{3\}$. L'état 3 est absorbant, l'état 2 n'est plus visité après un certain temps. Dans la terminologie qui est développée plus loin, les classes $\{0,1\}$ et $\{3\}$ sont *récurrentes,* la classe $\{2\}$ est *transiente.*

5.6 ÉTATS RÉCURRENTS ET TRANSIENTS

Pour tout état j, désignons par T_j le *temps d'atteinte* de la chaîne (X_n) $(n \geq 0)$ dans l'état j à partir de l'instant 1. Autrement dit,

$$T_j := \inf\{n \geq 1 : X_n = j\}. \tag{5.6.1}$$

Ce temps d'atteinte est un *temps d'arrêt* de la chaîne (*cf.* chap. 2). Rappelons que cela signifie que pour tout $n \geq 0$, l'évènement $\{T_j = n\}$, qui est égal à $\{X_1 \neq j, \ldots, X_{n-1} \neq j, X_n = j\}$ ne dépend que de $X_1, \ldots, X_n$, ce qui est vrai.

Définition 5.6.1 (Loi de probabilité conditionnelle du temps d'atteinte). Pour tout couple (i,j) d'états et tout $n \geq 1$, on pose :

$$f_{i,j}^{(n)} := \mathrm{P}\{T_j = n \,|\, X_0 = i\}. \tag{5.6.2}$$

Ainsi, $f_{i,j}^{(n)}$ $(n \geq 1)$ est la probabilité pour que le processus, partant de l'état i, atteigne l'état j, *pour la première fois*, à l'instant n. Pour tout couple d'états (i,j), on pose, par convention, $f_{i,j}^{(0)} := 0$

Théorème 5.6.1. *Pour tout entier $n \geq 1$, on a l'identité :*

$$p_{i,j}^{(n)} = \sum_{k=0}^{n} f_{i,j}^{(k)}\, p_{j,j}^{(n-k)}. \tag{5.6.3}$$

Puisque $p_{i,j}^{(0)} = \delta_{i,j}$ (symbole de Kronecker, qui vaut 1 si $i = j$ et 0 autrement), cette identité est encore valable pour $n = 0$ et $i \neq j$.

Démonstration. Le processus passe de i à j en n étapes si et seulement s'il passe de i à j *pour la première fois en k étapes* $(0 \leq k \leq n)$ et s'il passe ensuite de j à j en les $(n-k)$ étapes suivantes. Ces chemins, pour des k distincts, sont disjoints et la probabilité d'un chemin pour k fixé est $f_{i,j}^{(k)}\, p_{j,j}^{(n-k)}$.

Ce raisonnement intuitif peut être rendu rigoureux de la façon suivante. Comme, pour $n \geq 1$,

$$\{X_n = j\} = \sum_{k=1}^{n-1} \{T_j = k, X_n = j\} + \{T_j = n\},$$

on en déduit

$$\begin{aligned}
\mathrm{P}\{X_n = j \,|\, X_0 = i\} &= \sum_{k=1}^{n-1} \mathrm{P}\{T_j = k, X_n = j \,|\, X_0 = i\} + f_{i,j}^{(n)} \\
&= \sum_{k=1}^{n-1} \mathrm{P}\{T_j = k \,|\, X_0 = i\}\, \mathrm{P}\{X_n = j \,|\, T_j = k, X_0 = i\} \\
&\qquad + f_{i,j}^{(n)}.
\end{aligned}$$

Or, pour $1 \leq k \leq n-1$, l'évènement $\{T_j = k, X_0 = i\}$ est de la forme $\{A, X_k = j\}$, où A appartient à $\mathfrak{A}_{k-1} \subset \mathfrak{A}_k$. Par conséquent,

$$\mathrm{P}\{X_n = j \,|\, T_j = k, X_0 = i\} = \mathrm{P}\{X_n = j \,|\, A, X_k = j\} = p_{j,j}^{(n-k)},$$

d'après (5.4.4). Comme $\mathrm{P}\{X_n = j \,|\, X_0 = i\} = p_{i,j}^{(n)}$ et $\mathrm{P}\{T_j = k \,|\, X_0 = i\} = f_{i,j}^{(k)}$, il en résulte

$$p_{i,j}^{(n)} = \sum_{k=1}^{n-1} f_{i,j}^{(k)}\, p_{j,j}^{(n-k)} + f_{i,j}^{(n)} = \sum_{k=0}^{n} f_{i,j}^{(k)}\, p_{j,j}^{(n-k)}. \quad \square$$

L'identité (5.6.3) permet de déterminer les $f_{i,j}^{(n)}$ par récurrence à partir des $p_{i,j}^{(n)}$:

$$\begin{aligned} f_{i,j}^{(1)} &= p_{i,j}^{(1)}; \\ f_{i,j}^{(n)} &= p_{i,j}^{(n)} - \sum_{k=1}^{n-1} f_{i,j}^{(k)}\, p_{j,j}^{(n-k)} \quad (n \geq 2). \end{aligned} \tag{5.6.4}$$

Posons

$$f_{i,j} := \mathrm{P}\{T_j < +\infty \mid X_0 = i\} = \sum_{n\geq 1} f_{i,j}^{(n)}. \tag{5.6.5}$$

C'est la probabilité pour que le processus, partant de i, passe par j *au moins une fois* au cours du temps ; si $i = j$, le nombre $f_{j,j}$ est la probabilité pour que le processus, partant de j, retourne en j, au moins une fois au cours du temps.

Définition 5.6.2. On dit que l'état j est *récurrent*, si $f_{j,j} = 1$. On dit qu'il est *transient* ou *transitoire*, si $f_{j,j} < 1$.

Théorème 5.6.2 (Critères de récurrence). *Un état j est récurrent ou transient selon que*

$$\sum_{n\geq 0} p_{j,j}^{(n)} = +\infty, \tag{5.6.6}$$

ou que

$$\sum_{n\geq 0} p_{j,j}^{(n)} < +\infty. \tag{5.6.7}$$

Remarque. Ces formules ont une interprétation intuitive : notons $N_j := \sum_{n\geq 0} I_{\{X_n = j\}}$ le nombre de *retours* dans l'état j après l'instant 0. Alors le nombre moyen $\mathbb{E}[N_j] = \sum_{n\geq 0} p_{j,j}^{(n)}$ est infini si et seulement si j est récurrent.

Il y a plusieurs techniques pour démontrer ce théorème. Nous utilisons ici la technique des *fonctions génératrices*, en faisant usage du Lemme d'Abel, dont nous rappelons les deux formes (*cf.*, par exemple, [FF1], chap. 9, lemme de la Proposition 2.3).

Lemme 5.6.3 (Lemme d'Abel).

(1) *Si la série de terme général α_n converge et a pour somme α, alors* $\lim_{s\to 1-0} \sum_{n\geq 0} \alpha_n s^n = \alpha$.

(2) *Si les α_n sont positifs et si $\lim_{s\to 1-0} \sum_{n\geq 0} \alpha_n s^n = \alpha \leq +\infty$, alors la série de terme général α_n a pour somme α.*

Afin de démontrer le Théorème 5.6.2, nous considérons les fonctions génératrices

$$P_{i,j}(s) := \sum_{n\geq 0} p_{i,j}^{(n)} s^n, \qquad F_{i,j}(s) := \sum_{n\geq 0} f_{i,j}^{(n)} s^n, \tag{5.6.8}$$

et à l'aide de la relation de récurrence (5.6.3) du Théorème 5.6.1, établissons une relation fonctionnelle entre elles.

On a :

$$\begin{aligned} P_{j,j}(s) &= 1 + \sum_{n\geq 1} p_{j,j}^{(n)}\, s^n = 1 + \sum_{n\geq 1} s^n \sum_{0\leq k\leq n} f_{j,j}^{(k)}\, p_{j,j}^{(n-k)} \\ &= 1 + \sum_{n\geq 0} s^n \sum_{0\leq k\leq n} f_{j,j}^{(k)}\, p_{j,j}^{(n-k)} \qquad \text{[puisque } f_{j,j}^{(0)} = 0] \\ &= 1 + F_{j,j}(s)\, P_{j,j}(s). \end{aligned}$$

Lorsque $i \neq j$, il vient de même : $P_{i,j}(s) = F_{i,j}(s)\, P_{j,j}(s)$. Se rappelant que $p_{i,j}^{(0)} = \delta_{i,j}$, la proposition suivante est démontrée.

Proposition 5.6.4. *On a les identités*

$$P_{j,j}(s) = \frac{1}{1 - F_{j,j}(s)}, \quad P_{i,j}(s) = F_{i,j}(s)\, P_{j,j}(s) \quad (i \neq j), \tag{5.6.9}$$

que l'on peut réunir en une seule formule :

$$P_{i,j}(s) = \delta_{i,j} + F_{i,j}(s)\, P_{j,j}(s). \tag{5.6.10}$$

Le Théorème 5.6.2 se démontre alors ainsi : supposons j récurrent, soit $\sum_{n\geq 0} f_{j,j}^{(n)} = 1$. D'après le Lemme d'Abel (1), $\lim_{s\to 1-0} \sum_{n\geq 0} f_{j,j}^{(n)}\, s^n = 1$ et donc $\lim_{s\to 1-0} F_{j,j}(s) = 1$. D'après (5.6.9), il en résulte $\lim_{s\to 1-0} P_{j,j}(s) = +\infty = \lim_{s\to 1-0} \sum_{n\geq 0} p_{j,j}^{(n)}\, s^n = +\infty$. On peut appliquer alors la partie (2) du Lemme d'Abel et conclure que $\sum_{n\geq 0} p_{j,j}^{(n)} = +\infty$.

Réciproquement, si j est transient, soit $\sum_{n\geq 0} f_{j,j}^{(n)} < 1$, suivant la même technique, on a $\lim_{s\to 1-0} P_{j,j}(s) < +\infty$, d'où, par le Lemme d'Abel (2), $\sum_{n\geq 0} p_{j,j}^{(n)} < +\infty$. □

Revenons à l'Exemple 2, § 5.5. Le calcul des coefficients diagonaux des matrices $\mathcal{P}^n$ est facile. Pour tout $n \geq 0$, on trouve : $p_{0,0}^{(n)} = p_{1,1}^{(n)} = 1/2$, $p_{2,2}^{(n)} = (1/4)^n$ et $p_{3,3}^{(n)} = 1$. Les séries $\sum p_{0,0}^{(n)}$, $\sum p_{1,1}^{(n)}$ et $\sum p_{3,3}^{(n)}$ divergent. Les états 0, 1, 3 sont récurrents. La série $\sum p_{2,2}^{(n)}$ converge, l'état 2 est transient.

Proposition 5.6.5. *Tout état de non-retour est transient ; tout état absorbant est récurrent.*

Démonstration. Pour un état j de non-retour, on a $p_{j,j}^{(n)} = 1$ si $n = 0$ et 0 autrement. La série de terme général $p_{j,j}^{(n)}$ est évidemment convergente ! L'état est donc transient. Pour un état absorbant, tous les termes $p_{j,j}^{(n)}$ de la série valent 1. La série est divergente et l'état est récurrent. □

Proposition 5.6.6. *La récurrence est une propriété de classe : si $i \leftrightsquigarrow j$ et si i est récurrent, alors j est aussi récurrent.*

Démonstration. Comme $i \leftrightsquigarrow j$, on a $p_{i,j}^{(n_1)} > 0$ et $p_{j,i}^{(n_2)} > 0$ pour certains entiers n_1, n_2. De là,

$$\sum_n p_{j,j}^{(n_2+n+n_1)} \geq \sum_n p_{j,i}^{(n_2)} p_{i,i}^{(n)} p_{i,j}^{(n_1)} = p_{j,i}^{(n_2)} p_{i,j}^{(n_1)} \sum_n p_{i,i}^{(n)} = +\infty. \quad \square$$

On peut donc parler de *classe récurrente* et de *classe transiente.* En fait, on a une propriété plus forte que celle décrite dans la proposition précédente : il suffit que j soit seulement *accessible* d'un état i récurrent pour que j lui-même soit récurrent (*cf.* Proposition 5.7.4). Terminons ce paragraphe par l'énoncé de quelques propriétés sur les états transients et d'une propriété sur les chaînes finies.

Proposition 5.6.7. *Soit j un état transient. Alors, pour tout état i, on a :*

$$\sum_{n\geq 0} p_{i,j}^{(n)} = \delta_{i,j} + f_{i,j} \sum_{n\geq 0} p_{j,j}^{(n)}; \tag{5.6.11}$$

$$\sum_{n\geq 0} p_{j,j}^{(n)} = \frac{1}{1-f_{j,j}}; \tag{5.6.12}$$

$$\sum_{n\geq 1} p_{i,j}^{(n)} = \frac{f_{i,j}}{1-f_{j,j}} \quad (i \neq j). \tag{5.6.13}$$

En particulier, la série de terme général $p_{i,j}^{(n)}$ est convergente et $p_{i,j}^{(n)} \to 0$, lorsque n tend vers l'infini.

Démonstration. L'identité (5.6.11) découle de l'identité (5.6.10), appliquée pour $s \to 1-0$ et du Théorème 5.6.2. Comme la série de terme général $p_{j,j}^{(n)}$ converge vers $\beta := \sum_{n\geq 0} p_{j,j}^{(n)}$, on a $\lim_{s\to 1-} P_{j,j}(s) = \beta$, par le Lemme d'Abel.

La série de terme général $f_{i,j}^{(n)}$ a pour somme $f_{i,j} \leq 1$. On a donc aussi $\lim_{s\to 1-} F_{i,j}(s) = f_{i,j}$. D'où $\lim_{s\to 1-} P_{i,j}(s) = \delta_{i,j} + f_{i,j}\,\beta$ et $\sum_{n\geq 0} p_{i,j}^{(n)} = \delta_{i,j} + f_{i,j}\,\beta < +\infty$. Les identités (5.6.12) et (5.6.13) sont de banales conséquences de la première formule (5.6.11). $\square$

Proposition 5.6.8. *Si une chaîne de Markov a un nombre fini d'états, elle a au moins un état récurrent.*

Démonstration. Soit N le nombre d'états de la chaîne. Pour tout entier $n \geq 0$, on a $\sum_{1\leq i,j\leq N} p_{i,j}^{(n)} = N$, d'où $\sum_{n\geq 0} \sum_{1\leq i,j\leq N} p_{i,j}^{(n)} = +\infty$, ce qui, d'après la Proposition 5.6.7, serait une contradiction, si tous les états étaient transitoires. $\square$

Remarque. La proposition précédente n'est pas vraie si la chaîne a une *infinité* d'états. On peut avoir une chaîne, où *tous* les états sont transients. C'est le cas,

par exemple, de la chaîne ayant $E = \mathbb{Z}$ pour ensemble des états et $\mathcal{P} = (p_{i,j})$ où $p_{i,i+1} = 1$ et $p_{i,j} = 0$ si $j - i \neq 1$, pour matrice de transition.

5.7 LA PROPRIÉTÉ DE MARKOV FORTE

Récrivons l'identité (5.4.6), se rappelant que A appartient à la tribu $\mathfrak{A}_n = \mathfrak{T}(X_0, \dots, X_n)$; en supposant $\mathrm{P}\{A, X_n = i_n\} > 0$, on a :

$$\begin{aligned}\mathrm{P}\{X_{n+1} = j_{n+1}, \dots, X_{n+r} = j_{n+r} \mid A, X_n = i_n\} \\ = \mathrm{P}\{X_{n+1} = j_{n+1}, \dots, X_{n+r} = j_{n+r} \mid X_n = i_n\} \\ = p_{i_n, j_{n+1}}\, p_{j_{n+1}, j_{n+2}} \cdots p_{j_{n+r-1}, j_{n+r}}.\end{aligned} \tag{5.7.1}$$

Cette formule implique, par simple spécialisation des paramètres, à la fois les propriétés *markovienne* et *d'homogénéité* décrites dans le paragraphe 1. La propriété dite *de Markov forte* consiste à prolonger cette formule, lorsque le conditionnement des probabilités ne se fait plus par rapport à un présent *fixé* $\{X_n = i_n\}$, mais par rapport à un présent *aléatoirement fixé* $\{X_T = i\}$, où T est un *temps d'arrêt* (*cf.* chap. 2) de la chaîne de Markov.

Soit T un temps d'arrêt adapté à la suite (X_n) $(n \geq 0)$. Rappelons qu'un évènement A appartient à la tribu $\mathfrak{A}_T$, s'il a la propriété que, pour tout $n \geq 0$, l'intersection $A \cap \{T = n\}$ appartient à $\mathfrak{A}_n$ (*cf.* § 2.3).

Théorème 5.7.1 (Propriété de Markov forte). *Soit T un temps d'arrêt, à valeurs dans $[0, +\infty]$, adapté à une chaîne de Markov (X_n) $(n \geq 0)$, dont l'ensemble des états est E et la matrice de transition $\mathcal{P} = (p_{i,j})$ $((i,j) \in E^2)$. Soit A un évènement de la tribu $\mathfrak{A}_T$ tel que* $\mathrm{P}\{T < \infty, A, X_T = i\} > 0$. *Alors on a l'identité :*

$$\begin{aligned}\mathrm{P}\{X_{T+1} = j_1, \dots, X_{T+r} = j_r \mid T < \infty, A, X_T = i\} \\ = \mathrm{P}\{X_1 = j_1, \dots, X_r = j_r \mid X_0 = i\} \\ = p_{i,j_1}\, p_{j_1,j_2} \cdots p_{j_{r-1},j_r}.\end{aligned} \tag{5.7.2}$$

Démonstration. En effet, posons $A' := \{T < \infty, A, X_T = i\}$, $B := \{X_{T+1} = j_1, \dots, X_{T+r} = j_r\}$, puis $B_n := \{X_{n+1} = j_1, \dots, X_{n+r} = j_r\}$, pour $n \geq 0$. Il s'agit de prouver : $\mathrm{P}(B \mid A') = p_{i,j_1}\, p_{j_1,j_2} \cdots p_{j_{r-1},j_r}$ ou encore $\mathrm{P}(A'B) = p_{i,j_1}\, p_{j_1,j_2} \cdots p_{j_{r-1},j_r}\, \mathrm{P}(A')$.

D'abord,

$$\begin{aligned}\mathrm{P}(A'B) &= \sum_{n \geq 0} \mathrm{P}\{T = n, A, X_T = i, B\} = \sum_{n \geq 0} \mathrm{P}\{T = n, A, X_n = i, B_n\} \\ &= \sum_{n \geq 0} \mathrm{P}\{B_n \mid T = n, A, X_n = i\}\, \mathrm{P}\{T = n, A, X_n = i\}.\end{aligned}$$

En fait, n'interviennent dans la sommation ci-dessus que les termes tels que $\mathrm{P}\{T = n, A, X_n = i\} > 0$ et pour ceux-ci, on peut utiliser l'identité (5.7.1), l'évènement $\{T = n, A\}$ appartenant à $\mathfrak{A}_n$.

On a donc :

$$\begin{aligned}\mathrm{P}\{B_n \mid T=n, A, X_n=i\} &= \mathrm{P}\{B_n \mid X_n=i\}\\ &= \mathrm{P}\{X_{n+1}=j_1,\dots,X_{n+r}=j_r \mid X_n=i\} = p_{i,j_1}\,p_{j_1,j_2}\,\cdots\,p_{j_{r-1},j_r}.\end{aligned}$$

De là,

$$\begin{aligned}\mathrm{P}(A'B) &= p_{i,j_1}\,p_{j_1,j_2}\,\cdots\,p_{j_{r-1},j_r}\sum_{n\ge 0}\mathrm{P}\{T=n, A, X_n=i\}\\ &= p_{i,,j_1}\,p_{j_1,j_2}\,\cdots\,p_{j_{r-1},j_r}\mathrm{P}\{T<\infty, A, X_n=i\}.\end{aligned}$$

Par conséquent, $P(B \mid A') = p_{i,j_1}\,p_{j_1,j_2}\,\cdots\,p_{j_{r-1},j_r}$. □

La formule (5.7.2) est particulièrement utile, lorsque l'on prend pour temps d'arrêt le *temps d'atteinte* T_j de la chaîne dans l'état j (*cf.* (5.6.1)). Comme $X_{T_j}=j$, la formule (5.7.2) se récrit pour $i=j$:

$$\begin{aligned}\mathrm{P}\{X_{T_j+1}=j_1,\dots,X_{T_j+r}=j_r \mid T_j<\infty, A\}\\ &= \mathrm{P}\{X_1=j_1,\dots,X_r=j_r \mid X_0=j\}\\ &= p_{j,j_1}\,p_{j_1,j_2}\,\cdots\,p_{j_{r-1},j_r}. \qquad (5.7.3)\end{aligned}$$

On en tire le résultat suivant.

Corollaire 5.7.2. *Conditionnellement à l'évènement $\{T_j<\infty\}$, la suite translatée (X_{T_j+r}) $(r\ge 0)$ est une chaîne de Markov, de matrice de transition $\mathcal{P}$ et d'état initial j. De plus, la suite translatée (X_{T_j+r}) $(r\ge 1)$ est indépendante de la tribu $\mathfrak{A}_{T_j}$.*

Posons $\theta_{T_j}\{X_1=j_1,\dots,X_r=j_r\} := \{X_{T_j+1}=j_1,\dots,X_{T_j+r}=j_r\}$. L'opérateur θ_{T_j} est un opérateur de translation, qui fait subir à chacun des indices des X_i une translation égale à T_j. L'identité (5.7.3) se récrit, avec $A=\Omega$,

$$\begin{aligned}\mathrm{P}\{\theta_{T_j}\{X_1=j_1,\dots,X_r=j_r\} \mid T_j<\infty\}\\ &= \mathrm{P}\{X_1=j_1,\dots,X_r=j_r \mid X_0=j\}. \qquad (5.7.4)\end{aligned}$$

Si B est évènement appartenant à $\mathfrak{A}$, on peut donner un sens à $\theta_{T_j}\,B$ à l'aide du *processus canonique* décrit dans le paragraphe 5.2 et montrer que l'on a :

$$\mathrm{P}\{\theta_{T_j}\,B \mid T_j<\infty\} = \mathrm{P}\{B \mid X_0=j\}. \qquad (5.7.5)$$

Nous nous contentons d'utiliser (5.7.5) pour les seuls évènements $\{N_j=k\}$, où k est un entier positif, fini ou infini et où N_j est le *nombre de retours dans l'état* j après l'instant 0, à savoir

$$N_j = \sum_{n\ge 1} I_{\{X_n=j\}}. \qquad (5.7.6)$$

Le translaté $\theta_{T_j}\{N_j=k\}$ est obtenu en faisant subir à tous les évènements $\{X_n=j\}$ apparaissant dans la définition de N_j la translation θ_T. On a donc, avec

$$N'_j = \sum_{n\ge 1} I_{\{\theta_{T_j}\{X_n=j\}\}} = \sum_{n\ge 1} I_{\{X_{T_j+n}=j\}}, \qquad (5.7.7)$$

la relation :

$$\theta_{T_j}\{N_j = k\} = \{N'_j = k\}. \tag{5.7.8}$$

La propriété de Markov forte, dans la forme (5.7.5), permet donc d'écrire

$$\mathrm{P}\{N'_j = k \,|\, T_j < \infty\} = \mathrm{P}\{N_j = k \,|\, X_0 = j\}. \tag{5.7.9}$$

Or, N'_j est le nombre de retours en j, *après une premier retour* en j, soit $N_j = \sum_{1 \le n \le T_j} I_{\{X_n = j\}} + N'_j$. Comme il ne peut y avoir de retours en j *strictement* avant T_j, on a encore

$$N_j = \begin{cases} 1 + N'_j, & \text{si } T_j < \infty\,; \\ 0, & \text{sinon.} \end{cases}$$

D'où $\mathrm{P}\{N_j = k+1 \,|\, T_j < \infty\} = \mathrm{P}\{N_j = k \,|\, X_0 = j\} = \mathrm{P}^j\{N_j = k\}$ pour k entier positif, fini ou infini.

Prenons $\mu = \varepsilon_j$ comme loi de probabilité initiale (*cf.* § 5.4.1), de sorte que $\mathrm{P} = \mathrm{P}^j := \mathrm{P}^{\varepsilon_j}$. La probabilité $f_{j,j} = \mathrm{P}^j\{T_j < \infty\}$ a été considérée en (5.6.5). Comme $\{T_j < \infty\} = \{N_j \ge 1\}$, on en déduit, pour $0 \le k \le \infty$,

$$\begin{aligned} \mathrm{P}^j\{N_j = k+1\} &= \mathrm{P}^j\{N_j = k+1, N_j \ge 1\} \\ &= \mathrm{P}^j\{N_j = k+1, T_j < \infty\} \\ &= \mathrm{P}^j\{T_j < \infty\}\, \mathrm{P}^j\{N_j = k+1 \,|\, T_j < \infty\} \\ &= f_{j,j}\, \mathrm{P}^j\{N_j = k\}. \end{aligned}$$

Comme $\mathrm{P}^j\{N_j = 0\} = \mathrm{P}^j\{T_j = \infty\} = 1 - f_{j,j}$, deux cas se présentent :

(1) ou bien $f_{j,j} = 1$, l'état j est *récurrent* et alors $\mathrm{P}^j\{N_j = k\} = 0$ pour tout $k < \infty$, ce qui entraîne $\mathrm{P}^j\{N_j = \infty\} = 1$;

(2) ou bien $a := f_{j,j} < 1$, l'état j est *transient* et $\mathrm{P}^j\{N_j = k\} = a^k - a^{k+1}$ pour tout $k < \infty$. Cette dernière propriété entraîne que $\mathrm{P}^j\{N_j = \infty\} = 1 - \sum_{k<\infty} (1-a)a^k = 0$.

Ainsi

$$\mathrm{P}^j\{N_j = \infty\} = \begin{cases} 1, & \text{si } \mathrm{P}^j\{T_j < \infty\} = 1\,; \\ 0, & \text{si } \mathrm{P}^j\{T_j < \infty\} < 1. \end{cases} \tag{5.7.10}$$

On ne peut donc avoir $0 < \mathrm{P}^j\{N_j = \infty\} < 1$.

Enfin, comme $\mathbb{E}[N_j] = \mathbb{E}[\sum_{n \ge 1} I_{\{X_n = j\}}] = \sum_{n \ge 1} \mathrm{P}^j\{X_n = j\} = \sum_{n \ge 1} p_{jj}^{(n)}$, on retrouve bien les critères de récurrence du Théorème 5.6.2 sous la forme : $\mathbb{E}[N_j] = \infty$ si j est récurrent, tandis que $\mathbb{E}[N_j] = f_{j,j}/(1 - f_{j,j}) < \infty$ si j est transient. Exprimons ces résultats dans un théorème.

Théorème 5.7.3. *Dans une chaîne de Markov homogène, il n'y a que deux sortes d'états : les états récurrents et les états transients. Soient j un état transient et $a := f_{j,j} = \mathrm{P}^j\{T_j < \infty\}$ la probabilité d'au moins un retour de la*

chaîne dans l'état j. Alors $a < 1$ et le nombre aléatoire N_j de retours de la chaîne dans l'état j suit la loi géométrique :

$$\mathrm{P}^j\{N_j = k\} = (1-a)a^k \qquad (k \geq 0).$$

Terminons ce paragraphe par une extension de la Proposition 5.6.6 sur les états récurrents.

Proposition 5.7.4. *Soit j un état récurrent et $k \neq j$ tel que $j \rightsquigarrow k$, alors $k \rightsquigarrow j$, de sorte que k est aussi récurrent et dans la même classe que j. En particulier, une chaîne de Markov ne peut aller d'un état récurrent vers un état transient.*

Démonstration. Supposons que pour $n \geq 1$ on ait $p_{k,j}^{(n)} > 0$. Il s'agit de montrer que l'on a $p_{k,j}^{(m)} > 0$ pour un certain $m \geq 1$. Il suffit de démontrer la propriété pour $n = 1$, soit $p_{j,k} > 0$. Si l'on avait $\mathrm{P}\{X_m = j \,|\, X_0 = k\} = p_{k,j}^{(m)} = 0$ pour tout $m \geq 1$, on aurait $\mathrm{P}\{N_j = \infty \,|\, X_0 = k\} \leq \sum_{m \geq 1} p_{k,j}^{(m)} = 0$. De là, $\mathrm{P}\{N_j = \infty \,|\, X_0 = j\} = \sum_{l \neq k} p_{j,l}\, \mathrm{P}\{N_j = \infty \,|\, X_0 = l\} \leq \sum_{l \neq k} p_{j,l} = 1 - p_{j,k} < 1$ et j ne serait pas récurrent, contrairement à l'hypothèse. □

Soit j un état transient. D'après (5.7.10), avec une probabilité égale à 1, il existe un entier n_0 tel que le processus ne revienne plus en j après l'instant n_0. Si la chaîne de Markov a un nombre *fini* d'états, elle possède au moins un état récurrent, d'après la Proposition 5.6.8. On en déduit donc le résultat suivant.

Proposition 5.7.5. *Soit* Rec *l'ensemble des états récurrents d'une chaîne de Markov finie. Alors, avec une probabilité égale à 1, la chaîne se trouve dans* Rec *au bout d'un temps fini.*

5.8 PÉRIODICITÉ

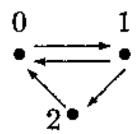
Il s'agit d'étudier dans quelles conditions le temps qui sépare deux retours au même état j est ou n'est pas multiple d'un temps minimum. On introduit pour ce faire la notion de *période*.

Définition 5.8.1. Soit j un état de retour ; on appelle *période* de j, le p.g.c.d. de tous les entiers $n \geq 1$ pour lesquels $p_{j,j}^{(n)} > 0$. On note $d(j)$ la période de j. Si $d(j) = d \geq 2$, on dit que j est *périodique de période* d ; si $d = 1$, on dit que j est *apériodique*. Si j est un état de *non-retour*, à savoir que, pour tout $n \geq 1$, on a $p_{j,j}^{(n)} = 0$ (voir § 5), on pose $d(j) = +\infty$.

Théorème 5.8.1. *Si i est périodique de période d finie et si $i \leftrightsquigarrow j$, $j \neq i$, alors j est aussi périodique de période d. La propriété de périodicité est une propriété de classe.*

Démonstration. Si $i \leftrightsquigarrow j$, alors il existe deux entiers n et m tels que $p_{i,j}^{(n)} > 0$ et $p_{j,i}^{(m)} > 0$. Comme i est de période $d(i) = d$, il existe aussi un entier $s \geq 1$ tel que $p_{i,i}^{(s)} > 0$. On a donc $p_{j,j}^{(m+s+n)} \geq p_{j,i}^{(m)} p_{i,i}^{(s)} p_{i,j}^{(n)} > 0$. Comme $p_{i,i}^{(s)} > 0 \Rightarrow p_{i,i}^{(2s)} > 0$, on a aussi : $p_{j,j}^{(m+2s+n)} > 0$. La période $d(j)$ de j divise donc à la fois $m+2s+n$ et $m+s+n$, donc aussi leur différence s, et en particulier la période $d(i)$ de i. De la même façon, on montre que $d(i)$ divise $d(j)$. Ainsi $d(j) = d(i) = d$. $\square$

Exemple 1. Considérons la chaîne de Markov, à trois états 0, 1, 2, dont le graphe associé est donné dans la figure ci-dessous, où toutes les flèches présentes correspondent à des probabilités de transition *strictement positives.*



L'état 0 est de retour ; les lacets $0 \to 1 \to 0$ et $0 \to 1 \to 2 \to 0$ ont pour longueurs 2 et 3, respectivement ; leur p.g.c.d. est $d = 1$; l'état 0 est donc apériodique. Maintenant, la chaîne est irréductible. Les deux autres états 1 et 2 sont aussi apériodiques, ce que l'on peut aussi vérifier directement.

Exemple 2. Dans la promenade sur $\mathbb{Z}$ décrite dans le paragraphe 2.3, tous les états sont périodiques, de période 2.

Exemple 3. Considérons la chaîne de Markov, dont l'ensemble des états est $E = \{0, 1, 2, 3\}$, la matrice de transition

$$\mathcal{P} = \begin{pmatrix} 0 & 0 & 1/2 & 1/2 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

et le graphe

0 1
3 2

Tous les états communiquent. Il y a donc une seule classe (récurrente). Il y a exactement deux lacets issus de 0 : $0 \to 2 \to 1 \to 0$ et $0 \to 3 \to 1 \to 0$, tous deux de longueur 3. La classe est donc périodique de période 3. Cette classe peut être subdivisée en *sous-classes cycliques* selon le schéma :

$$\{1\} \longrightarrow \{0\} \longrightarrow \{2,3\} \longrightarrow \{1\}$$

Si on ordonne les états dans l'ordre $1, 0, 2, 3$, la matrice de transition devient :

$$\begin{array}{c} \\ 1 \\ 0 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 0 & 2 & 3 \end{array} \\ \begin{pmatrix} 0 & \boxed{1} & 0 & 0 \\ 0 & 0 & \boxed{1/2} & \boxed{1/2} \\ \boxed{1} & 0 & 0 & 0 \\ \boxed{1} & 0 & 0 & 0 \end{pmatrix} \end{array}$$

Cet exemple rend intuitif le théorème suivant, donné sans démonstration.

Théorème 5.8.2. *Soit C une classe indécomposable, de période $d \geq 2$. Il existe une partition de C en sous-classes Γ_0, Γ_1, ... , Γ_{d-1} telles que les seules transitions possibles, dans la classe C, soient les transitions allant d'un état de Γ_i à un état de Γ_{i+1} pour $i = 0, 1, \ldots, d-1$, avec la convention que les indices i sont pris modulo d. Les sous-classes Γ_0, Γ_1, ... , Γ_{d-1} sont appelées les sous-classes cycliques de la classe indécomposable C.*

Ce théorème prend son importance, si la classe C est *récurrente*; dans ce cas, tout conséquent d'un état de C, situé dans Γ_i, est encore un état de C, situé dans Γ_{i+1}, les indices étant pris modulo d. Le passage de Γ_i à Γ_{i+1} est *obligatoire*, seul reste aléatoire le choix de l'état dans Γ_{i+1}.

5.9 COMPLÉMENTS ET EXERCICES

5.1 On considère la chaîne de Markov à deux états, décrite par la matrice de transition de l'Exemple 2.1 :

$$\mathcal{P} = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix} \qquad (0 < \alpha, \beta < 1).$$

(a) Écrire $\mathcal{P}$ sous la forme $\mathcal{P} = S\,D\,S^{-1}$, où D est la matrice diagonale $\mathrm{diag}(1, 1-\alpha-\beta)$ et S une matrice inversible.

(b) Pour tout $n \geq 0$, évaluer $\mathcal{P}^n$, puis $\lim_n \mathcal{P}^n$.

5.2 On reprend les mêmes notations que dans l'Exercice 5.1.

(a) Soit u une variable et I la matrice-unité. Évaluer le déterminant $\det(I - u\mathcal{P})$, puis la matrice inverse $(I - u\mathcal{P})^{-1}$.

(b) Développer en série la série de puissances, à coefficients matriciels : $\mathcal{P}(u) := \sum\limits_{n\geq 0} \mathcal{P}^n u^n = (1 - u\mathcal{P})^{-1}$. Retrouver l'expression de $\mathcal{P}^n$, puis de la valeur limite $\lim_n \mathcal{P}^n$.

À l'aide des deux exercices précédents, on peut donc calculer explicitement la matrice de transition en n étapes et déterminer la loi limite, lorsqu'il y a deux états. Celle-ci n'existe pas toujours lorsque le nombre d'états a est supérieur à 2. Le calcul de la matrice $\mathcal{P}^n$ peut devenir inextricable pour un nombre d'états grand. Déjà, pour $a = 3$, il y a de nombreux cas à considérer. Pour de grandes valeurs de a, les calculs ne sont possibles que si la matrice $\mathcal{P}$ a de nombreuses symétries ou beaucoup de zéros (matrice creuse). Le prochain exercice est consacré à l'étude d'une matrice très particulière d'ordre 3.

5.3 Considérons la matrice de transition

$$\mathcal{P} = \begin{pmatrix} 0 & 1-\alpha & \alpha \\ 1-\alpha & 0 & \alpha \\ 1/3 & 1/3 & 1/3 \end{pmatrix} \quad (0 < \alpha < 1).$$

(a) Vérifier que 1, $\alpha - 1$ et $\frac{1}{3} - \alpha$ sont valeurs propres de $\mathcal{P}$.

(b) On suppose $\alpha \neq \frac{2}{3}$, de sorte que les valeurs propres sont distinctes. Vérifier que les vecteurs $(1,1,1)$, $(1,-1,0)$, $(3\alpha, 3\alpha, -2)$ sont des vecteurs propres associés, respectivement, aux trois valeurs propres distinctes.

(c) On pose $S = \begin{pmatrix} 1 & 1 & 3\alpha \\ 1 & -1 & 3\alpha \\ 1 & 0 & -2 \end{pmatrix}$. La matrice inverse S^{-1} est donnée par :

$$S^{-1} = \frac{1}{2(2+3\alpha)} \begin{pmatrix} 2 & 2 & 6\alpha \\ 2+3\alpha & -(2+3\alpha) & 0 \\ 1 & 1 & -2 \end{pmatrix}.$$

(d) Présenter $\mathcal{P}^n$ comme un produit de trois matrices et en déduire sa valeur, puis celle de $\lim_n \mathcal{P}^n$. La première ligne u de cette matrice-limite forme une *loi de probabilité stationnaire* (étudiée au chapitre suivant) ; elle vérifie, en effet, l'identité : $u\,\mathcal{P} = u$.

5.4 Une fonction f de $E = \{1, 2, \ldots, n\}$ dans lui-même est dite *ultimement idempotente*, si l'on a $f^n = f^{n-1}$. Se reportant au paragraphe 2.6, caractériser $\lim_n \mathcal{P}_{f^n}$.

Dessiner le graphe associé à la matrice $\mathcal{P}_f$ avec f donné par

$$\begin{pmatrix} i \\ f(i) \end{pmatrix} = \begin{pmatrix} 1\ 2\ 3\ 4\ 5\ 6\ 7\ 8\ 9 \\ 5\ 1\ 1\ 9\ 7\ 6\ 7\ 2\ 6 \end{pmatrix}$$

et calculer $\lim_n \mathcal{P}_{f^n}$.

5.5 Soit (X_n) $(n \geq 0)$ une chaîne de Markov, de loi initiale μ_0 et de matrice de transition $\mathcal{P} = (p_{i,j})$ $(i,j) \in E^2)$.

(a) Pour $i \in E$ et $n \geq 0$, posons : $\mu_i^{(n)} := \mathrm{P}\{X_n = i\}$ et considérons le vecteur-ligne $\mu^{(n)} := (\mu_0^{(n)}, \mu_1^{(n)}, \ldots)$. Alors $\mu^{(n)} = \mu^{(0)}\,\mathcal{P}^n$.

(b) Pour $0 \leq m < n$ et $i, j \in E$, on a, avec les notations de (5.3.1), $\mathrm{P}\{X_m = i, X_n = j\} = \mu_i^{(m)}\, p_{i,j}^{(n-m)}$.

5.6 On reprend l'étude de la promenade au hasard sur $\mathbb{Z}$, décrite dans le paragraphe 2.3, où l'ensemble des états est $E = \mathbb{Z}$ et où, un nombre p tel que $0 < p < 1$ étant donné, la matrice de transition $\mathcal{P} = (p_{i,j})$ $((i,j) \in E^2)$ est telle que, pour tout $i \in \mathbb{Z}$, on a $p_{i,i+1} = 1 - p_{i,i-1} = p$ et $p_{i,j} = 0$ si $|i - j| \neq 1$.

(a) Ce processus a une seule classe indécomposable.

(b) Pour tout $n \geq 0$, on a : $p_{0,0}^{(2n+1)} = 0$ et $p_{0,0}^{(2n)} = \binom{2n}{n} p^n (1-p)^n$.

(c) On a $p_{0,0}^{(2n)} \sim (4p(1-p))^n / \sqrt{\pi\, n}$. [Utiliser la formule de Stirling.] La série de terme général $p_{0,0}^{(2n)}$ est convergente si et seulement si $p \neq \frac{1}{2}$. Donc la *promenade sur* $\mathbb{Z}$ *est récurrente si* $p = \frac{1}{2}$ *et transiente si* $p \neq \frac{1}{2}$.

(d) On a : $P_{0,0}(s) = \sum_{n \geq 0} p_{0,0}^{(2n)} s^{2n} = (1 - 4pqs^2)^{-1/2}$ $(q := 1-p)$.

(e) Notant $f_{0,0}^{(2n)}$ la probabilité pour que la promenade, partant de 0, retourne en 0, pour la première fois à l'instant $2n$, évaluer la fonction génératrice : $F_{0,0}(s) := \sum_{n \geq 1} f_{0,0}^{(2n)} s^{2n}$. Redémontrer le résultat de (c), à savoir que les états sont récurrents ou transients, si et seulement si $p = \frac{1}{2}$ ou $p \neq \frac{1}{2}$.

5.7 Promenade symétrique dans $\mathbb{Z}^d$ (Polyà)

On appelle ainsi le produit de d $(d \geq 1)$ promenades symétriques (chacune avec le paramètre $\frac{1}{2}$) dans $\mathbb{Z}$. Un retour en $(0, \ldots, 0) \in \mathbb{Z}^d$, à partir de $(0, \ldots, 0)$ équivaut à un retour simultané de toutes les d composantes de $0 \in \mathbb{Z}$ à $0 \in \mathbb{Z}$. Alors la promenade symétrique dans $\mathbb{Z}^d$ est récurrente si $d \leq 2$ et transiente si $d \geq 3$. En particulier, dans un réseau dans l'espace, si l'on part d'un point, il y a une probabilité non-nulle que l'on n'y revienne jamais !

5.8 Soit (Y_n) $(n \geq 0)$ une suite de variables aléatoires, indépendantes, identiquement distribuées, de loi commune $q\varepsilon_{-1} + p\varepsilon_{+1}$ $(0 < p < 1, p + q = 1)$. La suite de terme général $X_n := (Y_n + Y_{n+1})/2$ $(n \geq 0)$ *n*'est *pas* une chaîne de Markov.

5.9 Le prix d'un certain produit peut prendre les valeurs 1, 2, 3, 4, 5 selon les lois de l'offre et de la demande. Le prix X_n à la date n détermine la demande D_n à la date n, selon la relation $D_n = c - X_n$, où c est une constante supérieure à 5. L'offre C_n, à la date n, est donnée par $C_n = c - 3 + Y_n$, où (Y_n) est une suite de variables aléatoires, indépendantes, identiquement distribuées, de loi commune $\frac{1}{2}(\varepsilon_{-1} + \varepsilon_{+1})$. Les changements de prix se font selon les règles suivantes :

$$\begin{aligned} X_{n+1} - X_n &= +1, \quad \text{si } D_n - C_n > 0\,; \\ X_{n+1} - X_n &= -1, \quad \text{si } D_n - C_n < 0\,; \\ X_{n+1} - X_n &= 0, \quad\ \ \text{si } D_n - C_n = 0. \end{aligned}$$

(a) On fixe $X_0 = i_0$. Alors (X_n) $(n \geq 0)$ est une chaîne de Markov homogène dont l'ensembe des états est $E = \{1, 2, 3, 4, 5\}$.

(b) Déterminer la matrice de transition $\mathcal{P}$ et trouver la décomposition de l'ensemble des états en classes indécomposables.

Chapitre 6

Applications des chaînes de Markov

On donne dans ce chapitre d'autres propriétés sur les chaînes de Markov et la description de plusieurs exemples où ce modèle s'avère très fructueux, comme celui de la « ruine du joueur » et celui de la diffusion d'Ehrenfest. On aborde aussi l'étude des lois de probabilité stationnaires.

6.1 ENSEMBLES CLOS

On reprend les mêmes notations que dans le chapitre précédent, où l'on note (X_n) $(n \geq 0)$ une chaîne de Markov homogène, dont l'ensemble des états est E et la matrice de transition $\mathcal{P} = (p_{i,j})$ $((i,j) \in E^2)$.

Définition 6.1.1. Un ensemble C d'états est dit *clos* (ou *fermé*), si pour tout $i \in C$ et tout $j \notin C$, on a : $p_{i,j} = 0$.

Proposition 6.1.1. *Si C est un ensemble clos, alors aucun état n'appartenant pas à C n'est accessible à partir d'un état de C ; autrement dit, pour tout $i \in C$, tout $j \notin C$ et tout $n \geq 1$, on a $p_{i,j}^{(n)} = 0$. En particulier, pour tout $i \in C$, on a : $\sum\limits_{k \in C} p_{i,k} = 1$; ce qui montre que la restriction de la chaîne aux seuls états de C définit une chaîne de Markov homogène, ayant C comme ensemble d'états.*

Démonstration. Pour $n \geq 2$, $i \in C$ et $j \notin C$, on a :

$$p_{i,j}^{(n)} = \sum_{k \in C} p_{i,k}\, p_{k,j}^{(n-1)} + \sum_{k \notin C} p_{i,k}\, p_{k,j}^{(n-1)}.$$

Dans la première somme, $p_{k,j}^{(n-1)}$ est nul, par récurrence ; dans la seconde somme, $p_{i,k}$ l'est, par définition. D'où $p_{i,j}^{(n)} = 0$ et aussi

$$\sum_{l \in C} p_{i,l}^{(n)} = 1. \quad \square$$

Tout ensemble réduit à un état absorbant (*cf.* § 5.6) est clos. Tout ensemble d'états absorbants est aussi clos.

Proposition 6.1.2. *Tout ensemble clos est une réunion de classes indécomposables.*

Démonstration. Soit i un état d'un ensemble clos C. D'après la Proposition 6.1.1, s'il existe $n \geq 1$ et un état j tel que $p_{i,j}^{(n)} > 0$, alors $j \in C$. Donc C contient tous les états j tels que $i \rightsquigarrow j$ et par suite toute la classe indécomposable à laquelle appartient i. $\square$

Remarque 1. La réciproque n'est pas vraie : une réunion de classes indécomposables n'est *pas* nécessairement close. Par exemple, avec $E = \{0, 1, 2, \dots\}$ et le graphe associé $0 \xrightarrow{1} 1 \xrightarrow{1} 2 \xrightarrow{1} \cdots \xrightarrow{1} n \xrightarrow{1} \cdots$, chacun des états est transient et forme à lui seul une classe transiente. Aucune de ces classes n'est close.

En revanche, pour tout $n \geq 0$, l'ensemble $C_n := \{n, n+1, \dots\}$, réunion de classes transientes, est néanmoins clos.

Remarque 2. Une classe indécomposable close peut être ou récurrente ou transiente. Par exemple, dans la promenade au hasard sur $\mathbb{Z}$ (*cf.* § 5.2.3), il n'y a qu'une seule classe et elle est transiente.

Proposition 6.1.3. *Toute classe récurrente est close.*

Démonstration. Cette proposition résulte de la Proposition 5.7.4, où l'on a démontré que si i est récurrent et si $i \rightsquigarrow j$ avec $j \neq i$, alors $j \rightsquigarrow i$. Soit C la classe récurrente qui contient i et $j \notin C$. Si l'on avait $p_{i,j} > 0$, on aurait $i \rightsquigarrow j$ et donc $j \rightsquigarrow i$, par la Proposition 5.7.4. L'état j appartiendrait à C, une contradiction. $\square$

6.2 TEMPS D'ATTEINTE

Dans le paragraphe § 5.6, on a introduit le *temps d'atteinte* de la chaîne de Markov (X_n) $(n \geq 0)$ dans l'état j, comme étant la variable aléatoire

$$T_j := \inf\{n \geq 1 : X_n = j\}. \tag{6.2.1}$$

On a posé également $f_{i,j}^{(n)} := \mathrm{P}\{T_j = n \mid X_0 = i\}$, une probabilité que l'on note également $\mathrm{P}^i\{T_j = n\}$, en prenant, comme loi initiale de la chaîne, la loi de probabilité singulière ε_i. L'espérance mathématique de T_j par rapport à la loi P^i est notée $M_{i,j} := \mathbb{E}^i[T_j]$. C'est le *temps d'atteinte moyen dans* j à

partir de i. Rappelons que l'on a posé $f_{i,j} := \sum_{n\geq 1} f_{i,j}^{(n)}$, de sorte que si $f_{i,j} < 1$, il vient $M_{i,j} = +\infty$ et si $f_{i,j} = 1$, on a

$$\mathbb{E}^i[T_j] = \sum_{n\geq 1} n\, f_{i,j}^{(n)}, \tag{6.2.2}$$

une quantité qui peut être finie ou infinie. La quantité $M_{i,i}$ est appelée *temps de retour moyen dans* i ; l'inverse $1/M_{i,i}$ est la *fréquence moyenne de retour en* i. (Si $M_{i,i} = +\infty$, on pose $1/M_{i,i} := 0$.)

Définition 6.2.1. On dit que l'état i est *positif*, si $M_{i,i} < +\infty$ et qu'il est *nul* si $M_{i,i} = +\infty$.

Nous omettons la démonstration du théorème suivant.[1]

Théorème 6.2.1 (Critère de positivité). *Soit i un état ;*

si $\limsup_{n\to+\infty} p_{i,i}^{(n)} > 0$, *l'état i est positif ;*

si $\lim_{n\to+\infty} p_{i,i}^{(n)} = 0$, *l'état i est nul.*

Notons que si i est transient, on sait que $p_{i,i}^{(n)} \to 0$, d'après la Proposition 5.6.7. Donc tout état transient est nul.

Théorème 6.2.2. *La propriété de positivité (de nullité) est une propriété de classe.*

Démonstration. Ce théorème est une conséquence du Critère de positivité. En effet, si j est un état nul et si $j \leftrightsquigarrow i$ avec $i \neq j$, il existe des entiers n_1, n_2 tels que pour tout $n \geq 0$, on ait $p_{j,i}^{(n_2)} > 0$, $p_{i,j}^{(n_1)} > 0$ et $p_{j,j}^{(n_2+n+n_1)} \geq p_{j,i}^{(n_2)} p_{i,j}^{(n_1)} p_{i,i}^{(n)}$. Alors $\lim_n p_{j,j}^{(n)} = 0 \Rightarrow \lim_n p_{i,i}^{(n)} = 0$ et l'état i est nul aussi. □

Si $M_{i,i}$ est fini, on a forcément $f_{i,i} = 1$. Donc *tout état positif est récurrent.* De même, si $f_{i,i} < 1$, on a $M_{i,i} = +\infty$; d'où *tout état transient est nul.* Il y a donc des classes récurrentes positives, des classes récurrentes nulles et toutes les classes transientes sont nulles. En revanche, lorsque l'ensemble des états est *fini*, il n'y a pas de classe récurrente nulle, ou encore il y a *identité entre états positifs et récurrents*, d'une part, et *états nuls et transients*, d'autre part.

On peut vérifier cette dernière propriété de la façon suivante. Soit i un état récurrent et C la classe (forcément finie) qui le contient. Si C ne contient que i, on a $p_{i,i}^{(n)} = 1$, pour tout $n \geq 0$ et i est forcément positif. Si C est une classe récurrente *nulle* et si elle contient au moins deux états, prenons $j \in C$, $j \neq i$. Comme $p_{i,i}^{(n+m)} \geq p_{i,j}^{(n)} p_{j,i}^{(m)}$, en prenant m tel que $p_{j,i}^{(m)} > 0$, on en déduit $\lim_n p_{i,j}^{(n)} = 0$. De même $\lim_n p_{j,i}^{(n)} = 0$. Or $\sum_{i,j\in C} p_{i,j}^{(n)} = \operatorname{card} C$, d'où $\lim_n \sum_{i,j\in C} p_{i,j}^{(n)} = \operatorname{card} C > 0$, une contradiction. □

---

[1] Voir, par exemple, Loève (Michel). — *Probability Theory, I*, 4th Edition. — Springer-Verlag, New York, 1978, p. 33-36.

Théorème 6.2.3. *Les temps d'atteinte moyens $M_{i,j}$ vérifient dans $[1,+\infty]$ la relation :*

$$M_{i,j} = 1 + \sum_{k \neq j} p_{i,k}\, M_{k,j}. \tag{6.2.3}$$

Démonstration. On suppose que la chaîne est en i à l'instant 0 et on fait un conditionnement par rapport à la variable X_1. Alors $M_{i,j} = \mathbb{E}^i[T_j] = \mathbb{E}^i[\mathbb{E}^i[T_j \,|\, X_1]] = p_{i,j}\, \mathbb{E}^i[T_j \,|\, X_1 = j] + \sum_{k \neq j} p_{i,k}\, \mathbb{E}^i[T_j \,|\, X_1 = k] = p_{i,j} \cdot 1 + \sum_{k \neq j} p_{i,k}(1 + M_{k,j}) = 1 + \sum_{k \neq j} p_{i,k}\, M_{k,j}$. □

On peut récrire (6.2.3) sous forme matricielle comme suit : posons $M := (M_{i,j})$, puis notons M^δ la matrice obtenue de M en remplaçant les coefficients diagonaux par 0 et U la matrice dont tous les coefficients sont égaux à 1. Alors (6.2.3) se récrit :

$$M = U + \mathcal{P}\, M^\delta. \tag{6.2.4}$$

Si la chaîne de Markov est *irréductible* et si la seule classe dont elle est formée est *positive*, alors, par définition, tous les coefficients diagonaux $M_{i,i}$ sont *finis*. Notons Δ la matrice diagonale dont les coefficients diagonaux sont les $M_{i,i}$. On peut écrire $M^\delta = M - \Delta$, d'où

$$M = U + \mathcal{P}\,(M - \Delta). \tag{6.2.5}$$

Théorème 6.2.4. *Si la chaîne est irréductible et si la seule classe dont elle est formée est positive, alors la matrice M est à coefficients finis.*

Démonstration. En itérant (6.2.5), on obtient $M = U + \mathcal{P}M - \mathcal{P}\Delta = U + \mathcal{P}(U + \mathcal{P}M - \mathcal{P}\Delta) - \mathcal{P}\Delta$, d'où, puisque $\mathcal{P}U = U$, l'itération $M = 2U + \mathcal{P}^2 M - (\mathcal{P} + \mathcal{P}^2)\Delta = \cdots = nU + \mathcal{P}^n M - (\mathcal{P} + \cdots + \mathcal{P}^n)\Delta$. Comme $(\mathcal{P} + \cdots + \mathcal{P}^n)\Delta$ est une matrice à coefficients *finis*, on en déduit :

$$M + (\mathcal{P} + \cdots + \mathcal{P}^n)\Delta = nU + \mathcal{P}^n M. \tag{6.2.6}$$

En prenant le coefficient diagonal en (i,i), on obtient, pour tout $n \geq 1$, l'identité : $M_{i,i} + (p_{i,i} + \cdots + p_{i,i}^{(n)})M_{i,i} = n + \sum_k p_{i,k}^{(n)}\, M_{k,i}$, ou encore, pour tout $n \geq 1$,

$$(1 + p_{i,i} + \cdots + p_{i,i}^{(n)})M_{i,i} = n + \sum_k p_{i,k}^{(n)}\, M_{k,i}. \tag{6.2.7}$$

Soit alors j un état différent de i. Par hypothèse, $i \rightsquigarrow j$ et donc il existe $n_0 \geq 1$ tel que $\alpha := p_{i,j}^{(n_0)} > 0$. L'identité (6.2.7) pour $n = n_0$ donne : $+\infty > (1 + p_{i,i} + \cdots + p_{i,i}^{(n_0)})M_{i,i} = n_0 + \sum_k p_{i,k}^{(n_0)}\, M_{k,i} \geq n_0 + \alpha\, M_{j,i}$. D'où $M_{j,i}$ est fini et comme i et j sont arbitraires, le théorème est démontré. □

Corollaire 6.2.5. *Soit C une classe positive. Alors tous les coefficients $M_{i,j}$, pour i et j dans C, sont finis.*

Démonstration. En effet, la classe est forcément récurrente ; elle est donc close et la restriction de la chaîne aux seuls états de C est irréductible et formée évidemment d'une seule classe positive. □

Corollaire 6.2.6. *Si une chaîne est irréductible et a un nombre fini d'états, alors la matrice M est à coefficients finis.*

Démonstration. La seule classe indécomposable dont la chaîne est formée est nécessairement récurrente, donc positive. Le précédent théorème s'applique. □

Corollaire 6.2.7. *Si une chaîne est irréductible et a un nombre fini r d'états et si de plus sa matrice de transition $\mathcal{P}$ est bistochastique (i.e. la somme des coefficients par colonne vaut 1), alors les coefficients diagonaux $M_{i,i}$ de la matrice M sont tous égaux à r.*

Démonstration. En effet, $M_{i,j} = 1 + \sum_{k \neq j} p_{i,j}\, M_{k,j}$, d'où, en sommant sur i, la relation $\sum_i M_{i,j} = r + \sum_{k \neq j} \sum_i p_{i,j}\, M_{k,j} = r + \sum_{k \neq j} M_{k,j}$, d'où $M_{j,j} = r$ pour tout $j = 1, \ldots, r$. □

6.3 LE MODÈLE DE LA RUINE DU JOUEUR

Ce modèle a été décrit au paragraphe 5.2.4. On reprend les mêmes notations ; la fortune totale des deux joueurs est a et pour $n \geq 1$ on désigne par X_n la *fortune* du joueur A après la $n^{\text{ième}}$ partie. La matrice de transition $\mathcal{P} = (p_{i,j})$ $(0 \leq i, j \leq a)$ est donnée par : $p_{0,0} = p_{a,a} = 1$ et $p_{i,i+1} = p$, $p_{i,i-1} = q$ $(1 \leq i \leq a-1)$, les autres coefficients étant 0. La chaîne a deux classes récurrentes $\{0\}$, $\{a\}$ (états absorbants) et une classe transiente $\{1, 2, \ldots, a-1\}$.

Désignons par φ_i la probabilité pour que la chaîne, partant de i, atteigne l'état a *avant* l'état 0 (donc la probabilité de ruine pour le joueur B). Avec les notations des temps d'atteinte, $\varphi_i = \mathrm{P}\{T_a < T_0 \mid X_0 = i\}$, où

$$T_a = \inf\{n \geq 1 : X_n = a\}, \quad T_0 = \inf\{n \geq 1 : X_n = 0\}.$$

Posons :

$$T'_a = \inf\{n \geq 2 : X_n = a\}, \quad T'_0 = \inf\{n \geq 2 : X_n = 0\}.$$

D'abord,

$$\varphi_0 = 0, \quad \varphi_a = 1. \tag{6.3.1}$$

Le problème de la ruine est déjà résolu pour $a = 1$! Supposons $a \geq 2$ et $1 \leq i \leq a-1$. En conditionnant par rapport à l'état à la date 1, on voit que les nombres φ_i vérifient la récurrence :

$$\varphi_i = \varphi_{i-1}\, q + \varphi_{i+1}\, p \quad (1 \leq i \leq a-1). \tag{6.3.2}$$

Pour bien se convaincre que ce conditionnement est licite, on peut développer l'argument suivant. D'abord,

$$\mathrm{P}\{T_a < T_0, X_1 = 0 \mid X_0 = i\} = p_{i,0} = 0;$$
$$\mathrm{P}\{T_a < T_0, X_1 = a \mid X_0 = i\} = p_{i,a} = \begin{cases} p, & \text{si } i = a-1\,; \\ 0, & \text{autrement.} \end{cases}$$

Ensuite, pour $1 \le j \le a-1$, l'évènement $\{T_a < T_0, X_1 = j\}$ peut s'exprimer comme $\{1 + T'_a < 1 + T'_0, X_1 = j\}$. On en déduit, en utilisant les formules de conditionnement des chaînes de Markov (*cf.* § 5.4.3),

$$\begin{aligned}\mathrm{P}\{T_a < T_0, X_1 = j \mid X_0 = i\} &\\ &= \mathrm{P}\{1 + T'_a < 1 + T'_0, X_1 = j \mid X_0 = i\} \\ &= \mathrm{P}\{T'_a < T'_0 \mid X_1 = j, X_0 = i\}\,\mathrm{P}\{X_1 = j \mid X_0 = i\};\end{aligned}$$

d'où, puisque l'évènement $\{T'_a < T'_0\}$ ne dépend que de $X_2, X_3, \ldots$,

$$\begin{aligned}\mathrm{P}\{T_a < T_0, X_1 = j \mid X_0 = i\} &= \mathrm{P}\{T'_a < T'_0 \mid X_1 = j\}\, p_{i,j} \\ &= \mathrm{P}\{T_a < T_0 \mid X_0 = j\}\, p_{i,j} = \varphi_j\, p_{i,j}.\end{aligned}$$

Ainsi

$$\begin{aligned}\varphi_i = \mathrm{P}\{T_a < T_0 \mid X_0 = i\} &= \sum_{j=1}^{a} \mathrm{P}\{T_a < T_0, X_1 = j \mid X_0 = i\} \\ &= \sum_{j=1}^{a-1} \varphi_j\, p_{i,j} + p_{i,a} = \sum_{j=1}^{a} \varphi_j\, p_{i,j}.\end{aligned}$$

Soit la formule (6.3.2) annoncée : $\varphi_i = \varphi_{i-1}\, q + \varphi_{i+1}\, p \quad (1 \le i \le a-1)$.

Les formules (6.3.1) et (6.3.2) permettent de donner une forme explicite à la probabilité de ruine φ_i. En effet, pour $i = 1, \ldots, a-1$, on en déduit : $\varphi_{i+1} - \varphi_i = (q/p)(\varphi_i - \varphi_{i-1}) = (q/p)^2(\varphi_{i-1} - \varphi_{i-2}) = \cdots = (q/p)^i(\varphi_1 - \varphi_0)$. Comme $\varphi_0 = 0$, on obtient, pour $i = 2, \ldots, a$, la formule

$$\varphi_i - \varphi_1 = \varphi_1 \Big(\frac{q}{p} + \Big(\frac{q}{p}\Big)^2 + \cdots + \Big(\frac{q}{p}\Big)^{i-1} \Big). \tag{6.3.3}$$

D'où, encore pour $i = 2, \ldots, a$,

$$\varphi_i = \begin{cases} \dfrac{1 - \Big(\dfrac{q}{p}\Big)^i}{1 - \dfrac{q}{p}}\, \varphi_1, & \text{si } \dfrac{q}{p} \ne 1\,; \\[2ex] i\,\varphi_1, & \text{si } \dfrac{q}{p} = 1\,; \end{cases}$$

une formule qui, pour $i = a$, donne $1 = \varphi_a = (1 - (q/p)^a)/(1 - (q/p))\,\varphi_1$, soit $\varphi_1 = (1 - (q/p))/(1 - (q/p)^a)$, lorsque $p \ne \frac{1}{2}$ et $1 = a\varphi_1$, soit $\varphi_1 = 1/a$, lorsque

$p = \frac{1}{2}$. On en déduit la formule explicite, valable pour $i = 0, 1, \dots, a$,

$$\varphi_i = \begin{cases} \dfrac{1 - \left(\dfrac{q}{p}\right)^i}{1 - \left(\dfrac{q}{p}\right)^a}, & \text{si } p \neq \dfrac{1}{2}\,; \\[2ex] \dfrac{i}{a}, & \text{si } p = \dfrac{1}{2}. \end{cases} \tag{6.3.4}$$

On trouvera dans le chapitre 9 une autre technique pour aborder ce modèle, à l'aide de la théorie des martingales et du théorème d'arrêt.

6.4 LOIS DE PROBABILITÉ STATIONNAIRES

Dans ce paragraphe, il est commode de noter les lois de probabilité u sur l'ensemble dénombrable E comme des *vecteurs-lignes* $u = (u_0, u_1, \dots)$, où $u_i \geq 0$ et $\sum\limits_i u_i = 1$.

Définition 6.4.1. Soit u une loi de probabilité sur l'ensemble des états. Elle est dite *stationnaire*, si $u = u \cdot \mathcal{P}$, ou, de façon équivalente, si pour tout $j \in E$, on a :

$$u_j = \sum_{i \in E} u_i \, p_{i,j}. \tag{6.4.1}$$

Supposons qu'il existe une loi de probabilité stationnaire u ; pour tout $n \geq 0$, on a aussi : $u = u \cdot \mathcal{P}^n$. Il en résulte que si on prend u comme loi de probabilité *initiale* (*cf.* § 5.4.1), soit $\mu^{(0)} = u$, alors la loi de probabilité de X_n, qui est donnée par $\mu^{(n)} = \mu^{(0)} \mathcal{P}^n$, est *indépendante* de n et égale à $\mu^{(0)}$, soit $\mathrm{P}\{X_n = i\} = \mathrm{P}\{X_0 = i\} = \mu_i^{(0)}$.

De façon générale, si l'on prend u comme loi de probabilité initiale, la chaîne de Markov devient un *processus stationnaire*, c'est-à-dire, pour tout $n \geq 0$ et tout $k \geq 0$, on a : $\mathcal{L}(X_n, \dots, X_{n+k}) = \mathcal{L}(X_0, \dots, X_k)$. En effet, $\mathrm{P}\{X_n = i_0, \dots, X_{n+k} = i_k\} = \mathrm{P}\{X_n = i_0\}\,\mathrm{P}\{X_{n+1} = i_1, \dots, X_{n+k} = i_k \,|\, X_n = i_0\} = \mu_{i_0}^{(0)} p_{i_0,i_1} \cdots p_{i_{k-1},i_k} = \mathrm{P}\{X_0 = i_0\}\,\mathrm{P}\{X_1 = i_1, \dots, X_k = i_k \,|\, X_0 = i_0\} = \mathrm{P}\{X_0 = i_0, \dots, X_k = i_k\}$, d'après (5.4.7). D'où la terminologie.

Remarque 1. Il *n'existe pas nécessairement* de loi de probabilité stationnaire. Par exemple, avec $E = \mathbb{N}$ et le graphe associé

$$0 \xrightarrow{1} 1 \xrightarrow{1} 2 \xrightarrow{1} \cdots \xrightarrow{1} i \xrightarrow{1} \cdots,$$

le système $u_j = \sum\limits_{i \geq 0} u_i \, p_{i,j}$ $(i \geq 0)$ se réduit à $u_0 = 0$, $u_j = u_{j-1}$ $(j \geq 1)$. Les u_j sont donc identiquement nuls et u n'est pas une loi de probabilité sur E.

Remarque 2. Il peut exister une *infinité* de lois de probabilité stationnaires. Par exemple, dans le modèle de la ruine du joueur (*cf.* § 6.3), on vérifie que, pour tout α tel que $0 < \alpha < 1$, la loi $u = (\alpha, 0, \ldots, 0, 1 - \alpha)$ est une loi de probabilité stationnaire.

Le problème de l'existence et de l'unicité des solutions du système (6.4.1) a été résolu dans le cas *fini*. Dans le cas infini dénombrable, la situation est plus complexe. Nous ne donnerons que quelques indications. Traitons tout d'abord le cas fini.

Dans ce dernier cas, le fait que la matrice de transition est à coefficients réels *positifs* joue un rôle essentiel. Les résultats donnés ci-après reposent sur la théorie de Perron-Frobenius des *matrices à coefficients positifs.* Les résultats utiles pour la présente étude sont donnés dans les compléments de ce chapitre.

Rappelons qu'une suite (a_n) $(n \geq 1)$ de nombres réels *converge au sens de Cesaro* vers a, si la suite des moyennes $((a_1 + a_2 + \cdots + a_n)/n)$ $(n \geq 1)$ converge vers a, au sens ordinaire. On écrit indifféremment

$$a_n \xrightarrow{\mathrm{C}} a \quad \text{ou} \quad \lim_{n \to +\infty} {}_{\mathrm{C}}\, a_n = a.$$

Chacun doit avoir fait dans sa vie l'exercice qui consiste à montrer que si $a_n \to a$, alors $a_n \xrightarrow{\mathrm{C}} a$, la réciproque n'étant pas vraie. Pour s'en convaincre, on prend l'exemple traditionnel : $a_n = (-1)^n$.

Lemme 6.4.1. *Si le nombre des états est fini, alors* $\mathcal{P}^n \xrightarrow{\mathrm{C}} \Pi$. *De plus, la matrice* Π *est stochastique et l'on a :* $\Pi \mathcal{P} = \mathcal{P} \Pi = \Pi$ *et* $\Pi^2 = \Pi$.

Démonstration. Soit E l'ensemble (fini) des états de la chaîne. La relation $\mathcal{P}^n \xrightarrow{\mathrm{C}} \Pi$ signifie que pour tout couple d'états (i,j) on a : $(p_{i,j} + p_{i,j}^{(2)} + \cdots + p_{i,j}^{(n)})/n \to \pi_{i,j}$, où $\Pi = (\pi_{i,j})$ $((i,j) \in E^2)$. Ce résultat fait l'objet du Théorème $C\,4$, traité dans les compléments.

Maintenant, pour tout $i \in E$, on a : $\sum_{j \in E} \pi_{i,j} = \sum_{j \in E} \lim_n (p_{i,j} + p_{i,j}^{(2)} + \cdots + p_{i,j}^{(n)})/n = \lim_n \sum_{j \in E} (p_{i,j} + p_{i,j}^{(2)} + \cdots + p_{i,j}^{(n)})/n = \lim_n 1 = 1$. Ensuite, $\mathcal{P}^{n+1} = \mathcal{P}^n \cdot \mathcal{P} = \mathcal{P} \cdot \mathcal{P}^n$. Le résultat s'en déduit en prenant la limite de ces deux expressions au sens de Cesaro. □

Lemme 6.4.2. *Si l'état* j *est transient, alors pour tout* $i \in E$, *on a* $p_{i,j}^{(n)} \xrightarrow{\mathrm{C}} 0$, *d'où* $\pi_{i,j} = 0$.

Démonstration. D'après la Proposition 5.6.7, on sait que $p_{i,j}^{(n)} \to 0$, d'où $p_{i,j}^{(n)} \xrightarrow{\mathrm{C}} 0$. □

Remarque. Le précédent lemme est encore vrai lorsque E est infini dénombrable.

Lemme 6.4.3. *Supposons que la chaîne est irréductible et possède un nombre fini d'états. Alors, pour tout couple (i, j) d'états,*

$$\pi_{i,j} = \frac{1}{M_{j,j}} := \pi_j, \tag{6.4.2}$$

où $M_{j,j} = \mathbb{E}^j[T_j]$ est le temps de retour moyen de j en j.

Démonstration. D'après le Corollaire 6.2.5, la matrice $M = (M_{i,j})$ $((i,j) \in E^2)$ est à coefficients finis. Divisons les deux membres de l'identité $M + (\mathcal{P} + \cdots + \mathcal{P}^n)\,\Delta = n\,U + \mathcal{P}^n\,M$, établie dans le Théorème 6.2.4, par n et faisons tendre n vers l'infini. On obtient $\Pi\,\Delta = U$, soit, pour tout $(i,j) \in E^2$, la relation $\pi_{i,j}\,M_{j,j} = 1$, ou encore $\pi_{i,j} = 1/M_{j,j} =: \pi_j > 0$. □

Remarque. Ce lemme est encore vrai dans le cas infini, *dénombrable*, si le processus est *irréductible* et *positif.*

Théorème 6.4.4. *Si la chaîne de Markov a un ensemble fini d'états, alors*

(1) *(Existence) il existe au moins une loi de probabilité stationnaire ; de façon précise, à toute loi de probabilité initiale $\mu^{(0)}$, on peut associer une loi de probabilité stationnaire donnée par : $u = \mu^{(0)}\,\Pi$.*

(2) *(Unicité) les trois propriétés suivantes sont équivalentes :*

(a) *la loi de probabilité stationnaire est unique ;*

(b) *pour tout $(i,j) \in E^2$ on a : $\pi_{i,j} = \pi_j$;*

(c) *le processus n'admet qu'une seule classe récurrente.*

Si l'une de ces trois propriétés est vérifiée, l'unique loi de probabilité stationnaire est donnée par $u_j = \pi_j$ $(j \in E)$; en outre, cette loi de probabilité est portée par l'unique classe récurrente, c'est-à-dire, en désignant par $\mathcal{C}$ cette classe, on a les équivalences : $j \in \mathcal{C} \Leftrightarrow \pi_j > 0$ et $j \in \mathcal{C}^c \Leftrightarrow \pi_j = 0$.

Démonstration.

(1) Comme $\mathcal{P}^n \xrightarrow{\text{C}} \Pi$, d'après le Lemme 6.4.1, il suffit de vérifier que $u := \mu^{(0)}\,\Pi$ est solution de $u = u{\cdot}\mathcal{P}$. Or, $u{\cdot}\mathcal{P} = (\mu^{(0)}\,\Pi){\cdot}\mathcal{P} = \mu^{(0)}(\Pi\,\mathcal{P}) = \mu^{(0)}\Pi = u$.

(2) (a) $\Rightarrow$ (b). Supposons que la loi de probabilité stationnaire $u = \mu^{(0)}\Pi$ soit *unique*. Alors elle doit être indépendante de la loi de probabilité initiale $\mu^{(0)}$. En faisant successivement $\mu^{(0)} = (1, 0, \ldots, 0)$, ... , $\mu^{(0)} = (0, \ldots, 0, 1)$, on voit que l'on a $u_j = \pi_{1,j} = \pi_{2,j} = \cdots$, d'où (b).

(b) $\Rightarrow$ (a). Soit u une loi de probabilité stationnaire ; pour tout $n \geq 0$, elle vérifie $u = u\,\mathcal{P}^n$, d'où, en prenant la limite au sens de Cesaro, $u = u\,\Pi$. Utilisons maintenant l'hypothèse (b) : pour tout $j \in E$, on a : $u_j = \sum_{i \in E} u_i\,\pi_{i,j} = \sum_{i \in E} u_i\,\pi_j = \Big(\sum_{i \in E} u_i\Big)\pi_j = \pi_j$. La loi de probabilité u est unique et pour tout $j \in E$ on a : $u_j = \pi_j$.

(c) $\Rightarrow$ (b). Soit $\mathcal{C}$ l'unique classe récurrente et $\mathcal{T}$ la réunion des classes transientes. Trois cas sont à considérer :

(i) i arbitraire, $j \in \mathcal{T}$, alors $p_{i,j}^{(n)} \xrightarrow{\text{C}} = \pi_j$, d'après le Lemme 6.4.2.

(ii) $i, j \in \mathcal{C}$, alors $p_{i,j}^{(n)} \xrightarrow{C} \pi_j > 0$, car la restriction du processus à la classe $\mathcal{C}$ est *irréductible*. On peut donc faire usage du Lemme 6.4.3.

(iii) $i \in \mathcal{T}$, $j \in \mathcal{C}$, alors $p_{i,j}^{(n)} \xrightarrow{C} \pi_j > 0$, en faisant usage de l'argument suivant. Comme

$$p_{i,j}^{(m+k)} = \sum_{l \in \mathcal{C}} p_{i,l}^{(m)} p_{l,j}^{(k)} + \sum_{l \in \mathcal{T}} p_{i,l}^{(m)} p_{l,j}^{(k)},$$

on a :

$$\frac{1}{n} \sum_{k=1}^{n} p_{i,j}^{(m+k)} = \sum_{l \in \mathcal{C}} p_{i,l}^{(m)} \Big(\frac{1}{n} \sum_{k=1}^{n} p_{l,j}^{(k)} \Big) + \sum_{l \in \mathcal{T}} p_{i,l}^{(m)} \Big(\frac{1}{n} \sum_{k=1}^{n} p_{l,j}^{(k)} \Big).$$

Fixons m et faisons tendre n vers l'infini. D'après le Lemme 6.4.1, le membre de gauche de la précédente relation tend vers $\pi_{i,j}$ et la seconde expression entre parenthèses tend vers $\pi_{l,j}$. D'après le Lemme 6.4.3, la première expression entre parenthèses tend vers $\pi_j > 0$. Comme $\mathcal{C}$ et $\mathcal{T}$ sont des ensembles *finis*, donc aussi les sommes les concernant, on obtient :

$$\pi_{i,j} = \Big(\sum_{l \in \mathcal{C}} p_{i,l}^{(m)} \Big) \pi_j + \sum_{l \in \mathcal{T}} p_{i,l}^{(m)} \pi_{l,j}.$$

Faisons alors tendre m vers l'infini. Comme, pour tout $l \in \mathcal{T}$, on a $p_{i,l}^{(m)} \to 0$ et que les sommes qui interviennent sont sur des ensembles finis, il vient

$$\pi_{i,j} = \pi_j > 0.$$

(b) $\Rightarrow$ (c). Démontrons $\overline{(c)} \Rightarrow \overline{(b)}$. Supposons que le processus admette au moins deux classes récurrentes $\mathcal{C}_1$, $\mathcal{C}_2$. En numérotant les états dans l'ordre $\mathcal{C}_1$, $\mathcal{C}_2$, $\mathcal{R} := E \setminus (\mathcal{C}_1 \cup \mathcal{C}_2)$, les matrices $\mathcal{P}^n$ $(n \geq 0)$ et Π sont toutes de la forme :

$$\begin{array}{c} \\ \mathcal{C}_1 \\ \mathcal{C}_2 \\ \mathcal{R} \end{array} \begin{array}{c} \begin{array}{ccc} \mathcal{C}_1 & \mathcal{C}_2 & \mathcal{R} \end{array} \\ \begin{pmatrix} \square & 0 & 0 \\ 0 & \square & 0 \\ \square & \square & \square \end{pmatrix} \end{array}.$$

On voit que, pour la matrice Π, les lignes d'indice dans $\mathcal{C}_1$ ne correspondent pas avec celles d'indices dans $\mathcal{C}_2$. On a donc $\overline{(b)}$. $\square$

Théorème 6.4.5. *Considérons une chaîne de Markov irréductible, dont l'ensemble E des états est fini. Alors il existe une et une seule loi de probabilité stationnaire u donnée, pour tout $i \in E$, par :*

$$u_i = \pi_i = \frac{1}{M_{i,i}}. \tag{6.4.3}$$

Démonstration. Ce théorème résulte du Théorème 6.4.4 et du Lemme 6.4.3. On peut aussi en donner la démonstration directe suivante. Reprenons la formule (6.2.5), à savoir $M = U + \mathcal{P}(M - \Delta)$, dans laquelle la matrice $M = (M_{i,j})$ est à coefficients finis, dans laquelle Δ est la matrice diagonale $\text{diag}(M_{1,1}, M_{2,2}, \ldots)$ et U la matrice dont tous les coefficients sont égaux à 1.

On sait qu'*il existe* une loi de probabilité stationnaire u. Montrons qu'elle est *unique*. À cet effet, multiplions la précédente formule à gauche par u. On obtient $u\,M = u\,U + u\,\mathcal{P}(M - \Delta) = u\,U + u\,M - u\,\Delta$, soit $u\,\Delta = (u_1\,M_{1,1}, u_2\,M_{2,2}, \dots) = u\,U = (u_1 + u_2 + \cdots, u_1 + u_2 + \cdots, \cdots) = (1, 1, \dots)$, ou encore, pour tout $i \in E$, la relation $u_i = 1/M_{i,i}$. □

Remarque. On peut démontrer l'unicité d'une loi de probabilité stationnaire, de cette façon, dans le cas d'un processus à une infinité dénombrable d'états, à condition qu'il soit *irréductible* et *positif* (et non seulement récurrent).

6.5 INTERPRÉTATION DE LA STATIONNARITÉ

Introduisons les variables aléatoires suivantes :

$$N_n(j) := \sum_{k=1}^{n} I_{\{X_n=j\}};$$

(c'est le nombre de fois que le processus séjourne dans j dans $[1, n]$)

$$F_n(j) := \frac{1}{n} N_n(j);$$

(c'est la *fréquence* de séjour dans j dans l'intervalle $[1, n]$)

$$\mathbb{E}^i[I_{\{X_n=j\}}] = p_{i,j}^{(n)}.$$

D'où

$$\mathbb{E}^i[F_n(j)] = \frac{1}{n}(p_{i,j} + \cdots + p_{i,j}^{(n)});$$

c'est la *fréquence moyenne de séjour dans j durant l'intervalle de temps* $[1, n]$, sachant que le processus a débuté en i.

Le Lemme 6.4.1 s'interprète comme suit : pour tout couple $(i, j) \in E^2$, on a :

$$\mathbb{E}^i[F_n(j)] \to \pi_{i,j} \quad (n \to +\infty).$$

Ainsi, $\pi_{i,j}$ est la *fréquence asymptotique moyenne de séjour dans j*, sachant que le processus a débuté en i. La condition (2) (b) du Théorème 6.4.4 signifie que cette fréquence *ne dépend pas de l'état initial* dont le processus est parti. L'influence de l'état initial s'estompe au cours du temps.

La fin de ce paragraphe est consacrée à une étude partielle de la stationnarité, lorsque l'ensemble des états est *infini dénombrable*. Le Lemme 6.4.1 n'est plus valable. On peut seulement énoncer (sans démonstration) le lemme suivant.

Lemme 6.5.1. *On a $\mathcal{P}^n \xrightarrow{\text{C}} \Pi$, lorsque n tend vers l'infini. De plus Π est une matrice sous-stochastique, i.e. $\sum_j \pi_{i,j} \le 1$ pout tout i, et $\Pi = \Pi\,\mathcal{P} = \mathcal{P}\,\Pi$.*

Il peut même arriver que $\Pi = 0$, la matrice nulle. C'est le cas, par exemple, lorsque tous les états sont transients, puisque pour tout (i, j) on a $p_{i,j}^{(n)} \to 0$, d'où $\pi_{i,j} = 0$.

Théorème 6.5.2. *Si un processus n'a que des états transients, alors il n'existe pas de loi de probabilité stationnaire.*

Démonstration. Supposons qu'il existe u telle que $u\mathcal{P} = u$ et donc $u\mathcal{P}^n = u^n$. Alors, pour tout $n \geq 0$ et tout état j, on a : $u_j = \sum_i u_i p_{i,j}^{(n)}$. Comme tous les états sont transients, tous les termes $p_{i,j}^{(n)}$ tendent vers 0. Comme ensuite les termes $u_i p_{i,j}^{(n)}$ sont majorés par 1, on obtient : $u_j = \lim_n \sum_i u_i p_{i,j}^{(n)} = \sum_i \lim_n u_i p_{i,j}^{(n)} = 0$. Tous les u_j seraient nuls, ce qui contredirait $\sum_i u_i = 1$. □

Théorème 6.5.3. *Si un processus a au moins un état récurrent positif, alors il existe au moins une loi de probabilité stationnaire. En voici une explicite : soit i un état récurrent positif, donc satisfaisant $M_{i,i} = \mathbb{E}^i[T_i] < +\infty$. Pour tout état j, on pose*

$$\nu_j := \mathbb{E}^i\Big[\sum_{n\geq 0} I_{\{X_n=j,\, n\leq T_i-1\}}\Big]. \tag{6.5.1}$$

C'est le nombre moyen de visites dans l'état j entre deux visites dans l'état i. Alors $u = (u_j)$, avec $u_j := \nu_j/M_{i,i}$ $(j \in E)$, est une loi de probabilité stationnaire.

Démonstration. La relation (6.5.1) se récrit :

$$\begin{aligned}\nu_j &= \mathbb{E}^i\Big[\sum_{n\geq 0} I_{\{X_n=j,\, n\leq T_i-1\}}\Big] = \sum_{n\geq 0} \mathbb{E}^i[I_{\{X_n=j,\, n\leq T_i-1\}}] \\ &= \sum_{n\geq 0} \mathrm{P}^i\{X_n = j,\, n \leq T_i - 1\}.\end{aligned}$$

Si $j \neq i$, on a $\mathrm{P}^i\{X_0 = j, 0 \leq T_i - 1\} = 0$ et pour $n \geq 1$ on a aussi $\mathrm{P}^i\{X_n = j, n = T_i\} = 0$, de sorte que l'on a deux expressions pour ν_j :

$$\nu_j = \sum_{n\geq 1} \mathrm{P}^i\{X_n = j,\, n \leq T_i\}; \tag{6.5.2}$$

$$\nu_j = \sum_{n\geq 1} \mathrm{P}^i\{X_n = j,\, n \leq T_i - 1\}. \tag{6.5.3}$$

D'autre part, $\mathrm{P}^i\{X_0 = i,\, 0 \leq T_i - 1\} = 1$ et pour $n \geq 1$ toutes les probabilités $\mathrm{P}^i\{X_n = i,\, n \leq T_i - 1\} = 1$ sont nulles, de sorte que $\nu_i = 1$. Comme i est récurrent, la relation $1 = f_{i,i} = \sum_{n\geq 1} \mathrm{P}^i\{T_i = n\}$ est vérifiée ; or, elle peut encore s'écrire : $1 = \nu_1 = \sum_{n\geq 1} \mathrm{P}^i\{X_n = i,\, n \leq T_i\}$. Ainsi, (6.5.2) est vraie *pour tout* $j \in E$. On la récrit :

$$\begin{aligned}\nu_j &= \mathrm{P}^i\{X_1 = j,\, 1 \leq T_i\} + \sum_{n\geq 2} \mathrm{P}^i\{X_n = j,\, n \leq T_i\} \\ &= p_{i,j} + \sum_{n\geq 2} \mathrm{P}^i\{X_n = j,\, n \leq T_i\}\end{aligned}$$

Maintenant, pour $n \geq 2$,

$$\begin{aligned}\mathrm{P}^i\{X_n = j,\, n \leq T_i\} &= \sum_{k \neq i} \mathrm{P}^i\{X_{n-1} = k, X_n = j, n \leq T_i\} \\ &= \sum_{k \neq i} \mathrm{P}^i\{X_n = j \,|\, n \leq T_i, X_{n-1} = k\} \\ &\qquad\qquad \times \mathrm{P}^i\{X_{n-1} = k, n \leq T_i\}.\end{aligned}$$

Or, l'évènement $\{n \leq T_i\} = \{X_1 \neq i, \dots, X_{n-1} \neq i\}$ appartient à la tribu $\mathfrak{A}_{n-1}$. En outre, si k est différent de i, on a $\{n \leq T_i\} \cap \{X_{n-1} = k\} \neq \emptyset$, de sorte que l'on peut effacer $\{n \leq T_i\}$ du conditionnement du terme précédent. Ainsi

$$\begin{aligned}\mathrm{P}^i\{X_n = j,\, n \leq T_i\} &= \sum_{k \neq i} \mathrm{P}^i\{X_n = j \,|\, X_{n-1} = k\}\, \mathrm{P}^i\{X_{n-1} = k, n \leq T_i\} \\ &= \sum_{k \neq i} p_{k,j}\, \mathrm{P}^i\{X_{n-1} = k, n \leq T_i\}.\end{aligned}$$

Par suite,

$$\begin{aligned}\nu_j &= p_{i,j} + \sum_{n \geq 2} \mathrm{P}^i\{X_n = j,\, n \leq T_i\} \\ &= p_{i,j} + \sum_{n \geq 2} \sum_{k \neq i} p_{k,j}\, \mathrm{P}^i\{X_{n-1} = k, n \leq T_i\} \\ &= p_{i,j} + \sum_{k \neq i} p_{k,j} \sum_{n \geq 2} \mathrm{P}^i\{X_{n-1} = k, n \leq T_i\} \\ &= p_{i,j} + \sum_{k \neq i} p_{k,j} \sum_{n \geq 1} \mathrm{P}^i\{X_n = k, n \leq T_i - 1\} \\ &= p_{i,j} + \sum_{k \neq i} p_{k,j}\, p_{k,j} = \sum_k \nu_k\, p_{k,j}.\end{aligned}$$

D'autre part, dans $[0, +\infty]$, on a :

$$\begin{aligned}\sum_j \nu_j &= \sum_j \mathbb{E}^i\Big[\sum_{n \geq 0} I_{\{X_n = j, n \leq T_i - 1\}}\Big] = \mathbb{E}^i\Big[\sum_{n \geq 0} \sum_j I_{\{X_n = j, n \leq T_i - 1\}}\Big] \\ &= \mathbb{E}^i\Big[\sum_{n \geq 0} I_{\{n \leq T_i - 1\}}\Big] = \sum_{n \geq 0} \mathrm{P}^i\{T_i > n\} = \mathbb{E}^i[T_i] = M_{i,i}.\end{aligned}$$

Si i est récurrent *positif*, alors $M_{i,i} < +\infty$ et on peut poser $u_j := \nu_j / M_{i,i}$. Il en résulte que $u = (u_j)$ $(j \in E)$ est une loi de probabilité stationnaire. □

Théorème 6.5.4. *Si le processus a une infinité dénombrable d'états et s'il est irréductible et récurrent positif, alors il existe une et une seule loi de probabilité stationnaire* $u = (u_i)$ *donnée par* $u_i = 1/M_{i,i} > 0$ $(i \in E)$.

De plus, le nombre moyen ν_j *de visites dans l'état* j *entre deux visites dans l'état* i *est donné par :* $\nu_j = M_{i,i}/M_{j,j}$.

Démonstration. L'existence d'une telle loi a été établie dans le Théorème 6.5.3. Pour établir l'unicité, considérons une loi de probabilité stationnaire $u = (u_i)$ et utilisons, de nouveau, la relation (6.2.5), à savoir $M = U + \mathcal{P}(M - \Delta)$, qu'on multiplie à gauche par le vecteur-ligne u. On obtient : $u\,M = u\,U + (u\,\mathcal{P})(M - \Delta) = U + u(M - \Delta)$; soit $u\,\Delta = U$, ou encore, pour tout $i \in E$, la relation : $u_i\, M_{i,i} = 1$.

D'après le Théorème 6.5.3, pour tout $i \in E$, la loi $u = (u_j)$, avec $u_j = \nu_j / M_{i,i}$ $(j \in E)$, est une loi de probabilité stationnaire. D'après la première partie du présent théorème, on doit donc avoir $\nu_j / M_{i,i} = 1/M_{j,j}$, donc $\nu_j = M_{i,i}/M_{j,j}$. □

6.6 ERGODICITÉ

Le résultat principal, ci-dessous, est donné sans démonstration. En revanche, l'étude de l'ergodicité dans le modèle de diffusion d'Ehrenfest est faite très en détail.

Définition 6.6.1. On dit qu'un état est *ergodique*, s'il est récurrent positif et apériodique.

L'ergodicité est une propriété de classe, puisque c'est le cas pour la récurrence, la positivité et la périodicité. Une classe contenant un état ergodique est dite *classe ergodique*. Si l'ensemble des états est *fini*, si le processus est *irréductible* et si la seule classe dont il est formé est *apériodique*, alors cette classe est *ergodique*. On dit aussi que le processus est ergodique.

Théorème 6.6.1. *Si une chaîne de Markov a un nombre fini d'états et si toutes les classes récurrentes sont apériodiques (donc ergodiques), alors* $\mathcal{P}^n \to \Pi$, *ou encore, pour tout* $(i,j) \in E^2$, *on a :* $p_{i,j}^{(n)} \to \pi_{i,j}$. *De plus, la matrice* $\Pi = (\pi_{i,j})$ *est stochastique et vérifie :* $\Pi = \Pi\,\mathcal{P} = \mathcal{P}\,\Pi$.

Enfin, si le processus est irréductible et ergodique, pour tout couple $(i,j) \in E^2$, *on a :* $\pi_{i,j} = \pi_j > 0$.

Ce théorème, à rapprocher du Lemme 6.4.1, est donné sans démonstration. On peut, toutefois, l'établir en utilisant la formule (6.8.2) du Théorème 6.8.3 ci-après, en remarquant que si le processus est apériodique, la matrice de passage n'a pas de valeurs propres de module 1, différentes de 1.

Gardons les hypothèses de la première partie du Théorème 6.6.1 et désignons par $\mu^{(n)}$ la loi de X_n. Alors $\mu^{(n)} = \mu^{(0)}\,\mathcal{P}^n \to \mu^{(0)}\,\Pi =: u$, lorsque n tend vers l'infini et u est une loi de probabilité stationnaire. En d'autres termes, la loi de probabilité $\mu^{(n)}$ de X_n tend vers une loi de probabilité limite $u = \mu^{(0)}\,\Pi$, qui est stationnaire.

Reprenons maintenant les hypothèses de la seconde partie du Lemme. La loi limite $u = \mu^{(0)}\,\Pi = (\pi_1, \pi_2, \dots)$ ne dépend *pas* de la loi initiale. Si on prend cette loi limite comme loi initiale, le processus devient stationnaire. Il en résulte que l'évolution du processus au cours du temps est de plus en plus indiscernable d'une évolution stationnaire.

Le modèle de diffusion d'Ehrenfest

Ce modèle a été décrit au paragraphe 5.2.2. Nous conservons les mêmes notations. Rappelons que le processus possède $(a+1)$ états : 0, 1, ... , a; de plus, les coefficients $p_{i,j}$ de la matrice de transition $\mathcal{P}$ sont donnés par :

$$p_{k,k-1} = \frac{k}{a} \quad (1 \le k \le a)\,; \qquad p_{k,k+1} = 1 - \frac{k}{a} \quad (0 \le k \le a-1);$$
$$p_{k,l} = 0 \quad |k-l| \ne 1.$$

Tous les états communiquent : le processus est donc *irréductible*; les états 0 et a sont des *barrières réfléchissantes* (définition évidente !). Chaque état est périodique, de période 2. Le processus n'est donc pas ergodique. En revanche, le Théorème 6.4.5 permet d'affirmer que le processus admet une et une seule loi de probabilité stationnaire $u = (u_0, u_1, \dots, u_a)$ avec $u_i = 1/M_{i,i} = 1/\mathbb{E}^i[T_i]$. En fait, au lieu de calculer l'espérance mathématique des temps d'atteinte, on peut déterminer u directement à partir de sa définition $u = u\,\mathcal{P}$, c'est-à-dire résoudre le système

$$u_j = \sum_{i=0}^{a} u_i\, p_{i,j} \quad (0 \le j \le a),$$

dont on sait déjà qu'il a une solution unique, lorsque les u_i sont positifs et de somme 1. Ce système se récrit :

$$\begin{cases} u_0 = \dfrac{u_1}{a}, \quad u_a = \dfrac{u_{a-1}}{a}, \\ u_k = \Bigl(1 - \dfrac{k-1}{a}\Bigr) u_{k-1} + \dfrac{k+1}{a} u_{k+1} \quad (1 \le k \le a-1). \end{cases}$$

On calcule successivement $u_1 = a\,u_0$, $u_2 = (a(a-1)/2)\,u_0$, ce qui suggère la solution $u_k = \binom{a}{k} u_0$ $(0 \le k \le a)$. On vérifie immédiatement, par récurrence sur k, que c'est bien, en effet, la solution générale. Comme, par ailleurs, on doit avoir $1 = u_0 + u_1 + \cdots + u_a = u_0(\binom{a}{0} + \binom{a}{1} + \cdots + \binom{a}{a}) = u_0\, 2^a$, on trouve : $u_0 = 1/2^a$. La loi de probabilité stationnaire est donc donnée par $u_k = (1/2^a)\binom{a}{k}$ $(i = 0, 1, \dots, a)$; c'est la loi binomiale $B(a, \frac{1}{2})$. Pour $k = 0, 1, \dots, a$, on a : $M_{k,k} = \mathbb{E}^k[T_k] = 1/u_k = 2^a/\binom{a}{k}$. En particulier, $M_{0,0} = M_{a,a} = 2^a$.

On rappelle que, pour tout $n \ge 0$, la variable X_n est le nombre de boules dans l'urne A. Déterminons une relation de récurrence pour les nombres $m_n := \mathbb{E}[X_n]$ $(n \ge 0)$, une loi initiale ayant été fixée. Comme

$$\begin{aligned} \mathbb{E}[X_{n+1} - X_n \,|\, X_n = k] &= \mathbb{E}[X_{n+1} \,|\, X_n = k] - k \\ &= (k-1)\frac{k}{a} + (k+1)\Bigl(1 - \frac{k}{a}\Bigr) - k \\ &= 1 - 2\frac{k}{a}, \end{aligned}$$

on obtient

$$\mathbb{E}[X_{n+1} - X_n] = \sum_{k=0}^{a} \mathrm{P}\{X_n = k\}\Big(1 - 2\frac{k}{a}\Big)$$
$$= 1 - \frac{2}{a}\sum_{k=0}^{a} k\,\mathrm{P}\{X_n = k\} = 1 - \frac{2}{a}\,\mathbb{E}[X_n];$$

soit

$$m_{n+1} - m_n = 1 - \frac{2}{a}\,m_n.$$

Cette relation peut se récrire $m_{n+1} - a/2 = (1 - 2/a)(m_n - a/2)$, d'où

$$m_n - \frac{a}{2} = \Big(1 - \frac{2}{a}\Big)^n\Big(m_0 - \frac{a}{2}\Big);$$

ce qui montre que m_n tend vers $a/2$ exponentiellement, une limite qui est précisément l'espérance mathématique de la loi de probabilité stationnaire $B(a, \frac{1}{2})$. Avec ε_a comme loi initiale et donc $\mathbb{E}[X_0] = m_0 = a$, on obtient

$$m_n - \frac{a}{2} = \Big(1 - \frac{2}{a}\Big)^n \frac{a}{2}.$$

Dans ce cas et si $a > 2$, alors $m_n \downarrow a/2$, de façon exponentielle.

Ces résultats peuvent être interprétés comme suit : quel que soit le nombre de boules initialement contenues dans l'urne A, après un intervalle de temps assez long, la probabilité de trouver k boules dans A est approximativement la même que si les a boules avaient été réparties dans les urnes, chacune indépendamment avec une probabilité $\frac{1}{2}$. Pour n grand, on peut approcher la loi binomiale par la loi normale. Avec le temps, il y a, à peu près, autant de boules dans une urne que dans l'autre.

6.7 MATRICES BISTOCHASTIQUES

Une matrice stochastique $\mathcal{P}$ est dite *bistochastique* si sa transposée ${}^t\mathcal{P}$ est elle aussi stochastique. Dans une telle matrice, tous les coefficients sont donc positifs et les sommes par ligne *et* par colonne valent 1.

Si $\mathcal{P}$ est bistochastique et $n \geq 0$, alors $\mathcal{P}^n$ est aussi bistochastique et de même la matrice $(\mathcal{P} + \cdots + \mathcal{P}^n)/n$. Ces deux propriétés sont immédiates à vérifier. De même, il est banal de remarquer que toute matrice de permutation est bistochastique et par conséquent toute combinaison linéaire convexe d'un ensemble fini de matrices de permutations (de même ordre) est bistochastique. La réciproque est aussi vraie : toute matrice bistochastique est combinaison linéaire convexe de matrices de permutations, une propriété plus difficile à vérifier.

Théorème 6.7.1. *Considérons une chaîne de Markov homogène, dont l'ensemble des états est $E = \{1, 2, \ldots, r\}$ et la matrice de transition $\mathcal{P}$ est supposée bistochastique. Si la chaîne admet une seule classe récurrente, elle*

est nécessairement irréductible. Elle admet, en outre, une et une seule loi de probabilité stationnaire $u = (u_1, \ldots, u_r)$ donnée, pour tout $i = 1, \ldots, r$, par $u_i = 1/r$. En particulier, la fréquence asymptotique moyenne de séjour en i est $u_i = 1/r$ et le temps moyen de retour en i est $M_{i,i} = r$ $(1 \le i \le r)$.

Démonstration. D'après le Lemme 6.4.1, on a $\mathcal{P}^n \xrightarrow{\mathrm{C}} \Pi$, donc aussi ${}^t(\mathcal{P}^n) \xrightarrow{\mathrm{C}} {}^t\Pi$ et ${}^t\Pi$ est stochastique, donc Π est bistochastique. Comme la chaîne admet une seule classe récurrente, il résulte du Théorème 6.4.4 (2) que les coefficients $\pi_{i,j}$ de la matrice limite $\Pi = (\pi_{i,j})$ ont la propriété suivante : pour tout (i,j) on a : $\pi_{i,j} = \pi_j$. De plus, il existe une et une seule loi de probabilité stationnaire $u = (u_1, \ldots, u_r)$ donnée par $u_j = \pi_j$ $(1 \le j \le r)$.

Il reste à évaluer les nombres π_j. Comme Π est bistochastique, pour tout $j = 1, \ldots, r$, on a $1 = \sum_{i=1}^{r} \pi_{i,j} = \sum_{i=1}^{r} \pi_j = r\,\pi_j$, d'où $u_j = \pi_j = 1/r$.

Comme $\pi_j = 1/r > 0$, il résulte encore du Théorème 6.4.4 que *tous* les états sont *récurrents*. Il n'y a donc qu'une classe indécomposable, qui est récurrente et la chaîne est irréductible. □

Remarque. Si le nombre des états est fini, égal à r et si la matrice de transition $\mathcal{P}$ est bistochastique, la loi de probabilité uniforme $u = (1/r, \ldots, 1/r)$ est *toujours* une loi de probabilité stationnaire. En effet, pour tout j, on a $1/r = \sum_{i=1}^{r} (1/r)\, p_{i,j} = (1/r) \sum_{i=1}^{r} p_{i,j}$, puisque $\mathcal{P}$ est bistochastique. C'est le fait que la chaîne admet une seule classe récurrente qui permet de vérifier que cette loi de probabilité stationnaire est unique.

S'il y a au moins deux classes récurrentes, il n'y a plus unicité, comme le montre l'exemple suivant : $E = \{1,2\}$, $\mathcal{P} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. On voit que, pour tout α tel que $0 \le \alpha \le 1$, la loi de probabilité $u = (\alpha, 1-\alpha)$ est stationnaire.

6.8 COMPLÉMENTS SUR LES MATRICES POSITIVES

Soit $A = (a_{i,j})$ $(1 \le i,j \le r)$ une matrice carrée,[1] d'ordre $r \ge 1$, à coefficients réels. Elle admet r valeurs propres, réelles ou non, distinctes ou confondues. Soit s $(1 \le s \le r)$ le nombre de ses valeurs propres *distinctes*, notées λ_1, ... , λ_s. Pour chaque $i = 1, \ldots, s$, on désigne par α_i la *multiplicité* de λ_i, de sorte que $\alpha_1 + \cdots + \alpha_s = r$ et que chaque λ_i est un zéro d'ordre α_i du polynôme caractéristique $\det(I - \lambda A)$ de la matrice A. Nous ordonnons ces valeurs propres, chacune prise un nombre de fois égal à sa multiciplicité, comme suit : $\underbrace{\lambda_1, \ldots, \lambda_1}_{\alpha_1 \text{ fois}}, \ \ldots\ , \underbrace{\lambda_s, \ldots, \lambda_s}_{\alpha_s \text{ fois}}.$

[1] La théorie spectrale des matrices à coefficients réels positifs remonte à O. Perron (Über Matrizen, *Math. Annalen*, **64**, 1907, p. 248–263) et à G. Frobenius (Über Matrizen aus positiven Elementen, *Sitz. Ber. Akad. Wiss. Phys.-math. Klasse, Berlin*, 1908, p. 471–476, 1909, p. 514–518, 1912, p. 466–477) et est souvent appelée *Théorie de Perron-Frobenius.* Nous ne donnons ici que les éléments indispensables pour l'étude des matrices stochastiques.

La théorie de la réduction d'une matrice à la *forme de Jordan* montre qu'il existe une matrice *régulière* S, d'ordre r, telle que

$$S^{-1}\,A\,S = \begin{pmatrix} D_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & D_2 & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & D_s \end{pmatrix} := A',$$

où, ici, les $\mathbf{0}$ désignent des matrices nulles et où, pour chaque $i = 1, \dots, s$, la matrice D_i est une matrice triangulaire supérieure, d'ordre α_i, dont les coefficients diagonaux sont égaux à λ_i, de la forme

$$D_i = \begin{pmatrix} \lambda_i & * & * & \dots & * \\ 0 & \lambda_i & * & \dots & * \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \cdot & \cdot & \cdot & \dots & * \\ 0 & 0 & 0 & \dots & \lambda_i \end{pmatrix}.$$

Une telle matrice D_i peut encore s'écrire $D_i = \lambda_i\, I_{\alpha_i} + u_i$, où I_{α_i} est la matrice-unité d'ordre α_i et u_i est une matrice triangulaire supérieure *stricte*. Pour chaque $i = 1, \dots, s$, notons J_i (resp. u_i) la matrice, d'ordre r, déduite de A' en remplaçant tous les blocs diagonaux D_j par 0, sauf le bloc D_i, que l'on remplace par I_{α_i} (resp. par u_i).

Dans ces conditions, la matrice A' peut se récrire

$$A' = \sum_{i=1}^{s} (\lambda_i\, J_i + u_i). \tag{6.8.1}$$

Les matrices J_i, u_i $(1 \le i \le s)$ vérifient les relations :

$$J_i\, J_j = J_i\, u_j = u_i\, J_j = u_i\, u_j = 0 \quad (i \ne j); \qquad \text{[relations d'orthogonalité]}$$
$$J_i^2 = J_i, \quad J_i\, u_i = u_i\, J_i = u_i;$$
$$(u_i)^n = 0, \quad \text{dès que } n \ge \alpha_i. \qquad [u_i \text{ est nilpotente}]$$

De plus,

$$\sum_{i=1}^{s} J_i = I_r \quad \text{(matrice unité d'ordre } r\text{)}.$$

De (6.8.1) et des relations d'orthogonalité, on déduit, pour tout $n \ge 1$,

$$(A')^n = \sum_{i=1}^{s} (\lambda_i\, J_i + u_i)^n. \tag{6.8.2}$$

Si la valeur propre λ_i est non-nulle, on pose : $Q'_i(n) := \Big(J_i + \dfrac{u_i}{\lambda_i}\Big)^n$. À cause des relations précédentes, $Q'_i(n) = \sum\limits_{i=0}^{n} \binom{n}{k} J_i^{n-k} (u_i/\lambda_i)^k = \sum\limits_{i=0}^{n} \binom{n}{k} (u_i/\lambda_i)^k$. Comme $(u_i)^k = 0$ dès que $k \ge \alpha_i$, on voit que $Q'_i(n)$ est un *polynôme en* n, *à coefficients matriciels*, de degré au plus égal à $\alpha_i - 1 < r$.

Si une valeur propre est nulle, disons $\lambda_1 = 0$, alors $(\lambda_1 J_1 + u_1)^n = (u_1)^n$ est nul dès que $n \geq \alpha_1$. Par convention, on définit $Q'_1(n)$ comme étant le polynôme nul (à coefficients matriciels). Avec cette convention, pour tout $k \geq 1$, tout $l \geq 1$ et tout couple (i,j) tel que $1 \leq i,j \leq s$, on a :

$$Q'_i(k)\, Q'_j(l) = \delta_{i,j}\, Q'_i(k+l). \tag{6.8.3}$$

De plus, pour $n \geq r$,

$$(A')^n := \sum_{i=1}^{s} (\lambda_i)^n\, Q'_i(n). \tag{6.8.4}$$

Pour tout $i = 1,\ldots,s$, posons $Q_i(n) := S\, Q'_i(n)\, S^{-1}$. On a alors l'identité $A^n = (S\,A'\,S^{-1})^n = S\,(A')^n\,S^{-1} = \sum_{i=1}^{s} (\lambda_i)^n\, Q_i(n)$, qui est valable pour $n \geq r$. On a ainsi démontré le théorème suivant.

Théorème 6.8.1. *Soit A une matrice carrée réelle d'ordre r, soient $\lambda_1, \ldots, \lambda_s$ ses valeurs propres distinctes, de multiplicité $\alpha_1, \ldots, \alpha_s$, respectivement. Alors, pour tout $n \geq r$, on a l'identité*

$$(A)^n = \sum_{i=1}^{s} (\lambda_i)^n\, Q_i(n), \tag{6.8.5}$$

où, pour chaque $i = 1,\ldots,s$, la matrice $Q_i(n)$ est un polynôme en n, à coefficients matriciels, de degré au plus égal à $\alpha_i - 1 < r$ et où pour tout $k \geq 1$, tout $l \geq 1$ et tout couple (i,j) tel que $1 \leq i,j \leq s$, on a la relation

$$Q_i(k)\, Q_j(l) = \delta_{i,j}\, Q_i(k+l).$$

Appliquons le Théorème 6.8.1 lorsque A est une matrice stochastique $\mathcal{P}$.

Théorème 6.8.2. *Soit $\mathcal{P}$ une matrice stochastique d'ordre r. Alors*

(a) *toute valeur propre de $\mathcal{P}$ est de module au plus égal* 1 *;*

(b) *si λ_i est une valeur propre de module* 1*, alors le polynôme $Q_i(n)$, qui lui correspond dans* (6.8.5)*, se réduit à une matrice constante Q_i.*

Démonstration. Partons de la représentation

$$(A)^n = \sum_{k=1}^{s} (\lambda_k)^n\, Q_k(n),$$

où $A = \mathcal{P}$ et multiplions-la à gauche par $Q_i(1)$; on obtient : $Q_i(n)(\mathcal{P})^n = (\lambda_i)^n Q_i(n+1)$. Introduisons alors la norme $\|A\| := \sup_k \sum_l |a_{k,l}|$ d'une matrice $A = (a_{k,l})$ et utilisons l'inégalité sur les normes, valable pour deux matrices A, B : $\|A\,B\| \leq \|A\|\,\|B\|$. On obtient : $|\lambda_i|^n\, \|Q_i(n+1)\| \leq \|Q_i(1)\|\,\|\mathcal{P}^n\| = \|Q_i(1)\| = K < +\infty$, puisque $\|\mathcal{P}^n\| = 1$.

Il en résulte que, pour tout i, on a : $|\lambda_i| \leq 1$. D'autre part, si $|\lambda_i| = 1$, le degré du polynôme $Q_i(n)$ est forcément 0. □

Utilisons maintenant le fait que toute matrice stochastique $\mathcal{P}$ admet la valeur propre $\lambda_1 := 1$ et désignons par Π la matrice constante $\Pi := Q_1$ de la représentation (6.8.5). Nous obtenons le théorème suivant.

Théorème 6.8.3. *Soit $\mathcal{P}$ une matrice stochastique d'ordre r. Alors, pour tout $n \geq r$, on a :*

$$(\mathcal{P})^n = \Pi + \sum_j (\lambda_j)^n Q_j + \sum_k (\rho_k)^n Q_k(n), \tag{6.8.6}$$

où la première (resp. la seconde) sommation $\sum_j$ (resp. $\sum_k$) est étendue aux valeurs propres λ_j, de module 1, distinctes de **1** *(resp. aux valeurs propres ρ_k, non-nulles, de module strictement inférieur à 1). En outre, $\Pi^2 = \Pi$ et $Q_j^2 = Q_j$ pour tout j, i.e. Π et les Q_j sont des projecteurs.*

Théorème 6.8.4. *Soit $\mathcal{P}$ une matrice stochastique d'ordre r. Alors la suite $(\mathcal{P}^n)$ $(n \geq 0)$ converge, au sens de Cesaro, vers une matrice Π, d'ordre r, qui est un projecteur ($\Pi^2 = \Pi$), soit*

$$\mathcal{P}^n \xrightarrow{\mathrm{C}} \Pi \quad ou \quad \frac{\mathcal{P} + \mathcal{P}^2 + \cdots + \mathcal{P}^n}{n} \to \Pi,$$

lorsque n tend vers l'infini.

Démonstration. Dans la relation (6.8.6), on a $\lim_n (\rho_k)^n = 0$; de plus, $Q_k(n)$ est un polynôme en n ; de là, $\lim_n (\rho_k)^n Q_k(n) = 0$ et aussi $\lim_{n\to+\infty} {}_{\mathrm{C}} (\rho_k)^n Q_k(n) = 0$. Ensuite, chaque Q_j est une matrice constante et l'inégalité

$$\left|\frac{\lambda_j + \cdots + \lambda_j^n}{n}\right| \leq \frac{1}{n}\frac{2}{|1 - \lambda_j|}$$

montre que $\lim_{n\to+\infty} {}_{\mathrm{C}} (\lambda_j)^n Q_j = 0$. D'après (8.6), on obtient bien la relation : $\lim_{n\to+\infty} {}_{\mathrm{C}} (\mathcal{P})^n = \Pi$. $\square$

Remarque (Méthode pour déterminer la matrice Π). Soit $\mathcal{P}$ une matrice stochastique d'ordre r. Alors la matrice $A := \frac{1}{2}(I_r + \mathcal{P})$ est encore une matrice stochastique d'ordre r. En outre, si λ est valeur propre de $\mathcal{P}$ admettant X comme vecteur propre, alors $(1 + \lambda)/2$ est valeur propre de A admettant X comme vecteur propre. On peut donc réduire parallèlement les matrices $\mathcal{P}$ et A. En désignant par $\lambda_1 = 1, \lambda_2, \ldots, \lambda_s$ les valeurs propres *distinctes* de $\mathcal{P}$, on a, pour $n \geq r$:

$$\mathcal{P}^n = \Pi + \sum_{i=2}^{s} (\lambda_i)^r Q_i(n),$$

$$A^n = \Pi + \sum_{i=2}^{s} \Big(\frac{1 + \lambda_i}{2}\Big)^r Q_i(n).$$

On sait que $\mathcal{P}^n \xrightarrow{\mathrm{C}} \Pi$. Par ailleurs, pour $i = 2, \ldots, s$, on a $|\lambda_i| \leq 1$ et $\lambda_i \neq 1$, d'où $|(1 + \lambda_i)/2| < 1$. Par conséquent, $A^n \to \Pi$, au sens ordinaire. Ainsi, pour

calculer la limite *au sens de Cesaro* de $\mathcal{P}^n$, on forme $A := \frac{1}{2}(I_r + \mathcal{P})$ et on calcule la limite, *au sens ordinaire*, de A^n.

Le résultat $A^n \to \Pi$ s'exprime encore par : $(1/2^n) \sum_{k=0}^{n} \binom{n}{k} \mathcal{P}^k \to \Pi$.

6.9 COMPLÉMENTS ET EXERCICES

6.1 Le score du jeu du tennis[1]

Deux joueurs A et B décident de faire un «jeu» de tennis. Le «jeu» est une succession de «coups» joués entre les deux joueurs. Le score initial est $(0,0)$. On attribue 15 points au premier coup gagné par l'un des joueurs, puis 30 au second, 40 au troisième et 50 au quatrième, de sorte que le score passe de $(0,0)$ à $(15,0)$ ou à $(0,15)$, puis à $(30,0)$, ou à $(15,15)$, ou encore à $(0,30)$... Le joueur qui gagne le jeu est le joueur qui atteint 50 points à la réserve près que l'autre doit avoir au plus 30 points.

Par conséquent, lorsque le score est de $(40,30)$ (resp. $(30,40)$), deux possibilités se présentent :

ou bien A (resp. B) gagne le coup suivant et le score passe à $(50,30)$ (resp. passe à $(30,50)$) et A (resp. B) a gagné le jeu ;

ou bien B (resp. A) gagne le coup et on considére que l'on revient au score $(30,30)$ et les deux joueurs doivent continuer à jouer.

On suppose que les coups sont joués indépendamment et que A a la probabilité p et B la probabilité q de gagner un coup ($0 \le p \le 1$, $p + q = 1$). Le processus du jeu est assimilé à une chaîne de Markov homogène, dont les états sont les scores possibles. Le graphe associé est représenté dans la Figure 6.1, les flèches horizontales dirigées vers la droite, ainsi que l'unique flèche verticale dirigée vers le bas, portent la probabilité p ; les autres flèches portent la probabilité q. Le joueur A gagne le jeu lorsque le score atteint l'un des états $(50,0)$, $(50,15)$, $(50,30)$; le joueur B gagne en $(0,50)$, $(15,50)$ ou $(30,50)$. En convenant qu'il y a une probabilité égale à 1 de rester dans l'un de ces six états, lorsqu'on l'a atteint, ces six états sont des états *absorbants* de la chaîne. (Dans le graphe, il faudrait mettre une boucle de poids 1 autour de chacun d'eux.)

Les scores $a_0 = (30,50)$, $a_1 = (30,40)$, $a_2 = (30,30)$, $a_3 = (40,30)$, $a_4 = (50,30)$ forment un ensemble clos ; le déroulement de la chaîne à l'intérieur de cet ensemble est identique à une chaîne de Markov qui est celle du modèle de la *ruine du joueur*.

(a) On sait donc calculer la probabilité d'atteindre a_4 en partant d'un des états a_1, a_2, a_3 (*cf.* § 6.3). Donner le résultat.

(b) Trouver la probabilité que A gagne en arrivant au score $(50,0)$ (jeu blanc), au score $(50,15)$.

---

[1] Cet exercice a été proposé dans J.G. Kemeny & J.L. Snell (*Finite Markov Chains*. Princeton, D. Van Nostrand, 1960, § 7.2.) La solution donnée ici reprend celle qui nous avait été fournie par P. Cartier.

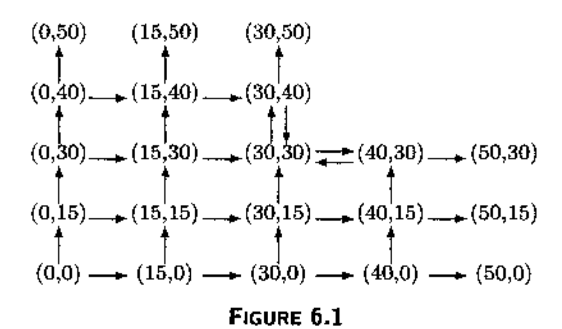


Figure 6.1

(c) Pour $i = 1, 2, 3$, soit T_i le temps d'atteinte de la chaîne dans l'état a_i. Déterminer les trois probabilités $p_i := \mathrm{P}^{(0,0)}\{T_i < \min\{T_j : j \neq i\}\}$ $(i = 1, 2, 3)$.

(d) La probabilité pour A de gagner le jeu vaut :

$$F(p) = \frac{15p^4 - 34p^5 + 28p^6 - 8p^7}{1 - 2p + 2p^2}.$$

Vérifier que $F(p) + F(1-p) = 1$ et en donner l'interprétation.

6.2 On considère une chaîne de Markov homogène, admettant $(r+1)$ états, numérotés de 0 à r $(r \geq 2)$. On suppose que si le processus est dans l'état 0, il a la probabilité $\frac{1}{2}$ d'y rester en une épreuve et la probabilité $1/(2r)$ de passer dans l'un quelconque des états 1, 2, ... , r. Il a, d'autre part, la probabilité nulle de passer à l'état 0 à partir de l'un quelconque des états 1, 2, ... , r. Enfin, s'il est dans l'état i $(1 \leq i \leq r-1)$ (resp. r), il a, en une épreuve, la possibilité d'y rester avec la probabilité $\frac{1}{2}$, ou la possibilité d'aller dans $(i+1)$ (resp. 1) avec la probabilité $\frac{1}{2}$.

(a) Donner la matrice de passage $\mathcal{P}$ pour $r = 2, 3$ et les différentes classes, ainsi que leur périodicité.

(b) Pour $r = 2$ et $n \geq 1$, déterminer $\mathcal{P}^n$; trouver $\lim_n \mathcal{P}^n$.

(c) Déterminer $f_{1,1}^{(n)} = \mathrm{P}\{T_1 = n \,|\, X_0 = 1\}$ (la probabilité que la chaîne retourne, pour la première fois, en 1, à la $n^{\text{ième}}$ transition). Évaluer l'espérance mathématique et la variance de T_1, par rapport à la mesure de probabilité $\mathrm{P}\{\cdot \,|\, X_0 = 1\}$.

(d) Pour $r \geq 2$ arbitraire, trouver $\lim_n \mathcal{P}^n$.

6.3 Mouvement cyclique

Une particule parcourt les sommets d'un polygone régulier à r côtés. Ces sommets sont numérotés $1, \ldots, r$, disons, dans le sens trigonométrique et constituent les r états du processus considéré ci-après. Si, à un instant déterminé, la particule est dans l'état i $(2 \leq i \leq r)$ (resp. dans l'état 1), alors, à l'instant suivant, elle sera dans l'état $(i+1)$ (resp. 2) avec la probabilité p

et dans l'état $(i-1)$ (resp. dans l'état r) avec la probabilité q. On suppose $0 < p < 1$ et $p + q = 1$. En notant X_n $(n \geq 0)$ l'état occupé par la particule à l'instant n, on définit ainsi une chaîne de Markov homogène, de matrice de transition

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ \vdots \\ r-2 \\ r-1 \\ r \end{array}\begin{array}{c} \begin{array}{ccccccc} 1 & 2 & 3 & \dots & r-2 & r-1 & r \end{array} \\ \begin{pmatrix} 0 & p & 0 & \dots & 0 & 0 & q \\ q & 0 & p & \dots & 0 & 0 & 0 \\ 0 & q & 0 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 0 & p & 0 \\ 0 & 0 & 0 & \dots & q & 0 & p \\ p & 0 & 0 & \dots & 0 & q & 0 \end{pmatrix} \end{array}.$$

Faire l'analyse de cette chaîne de Markov.

6.4 Séjour dans un labyrinthe

Un rat se déplace dans un labyrinthe, représenté dans la Figure 6.2, qui comporte neuf compartiments, numérotés de 1 à 9. À chaque unité de temps, il change de compartiment. Lorsqu'il est dans un compartiment ayant k portes $(2 \leq k \leq 4)$, il choisit l'une de ces k portes avec probabilité $1/k$. Le cheminement du rat est assimilé à une chaîne de Markov homogène dont l'ensemble des états est formé des neuf compartiments.

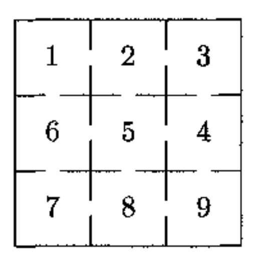


FIGURE 6.2

(a) Écrire la matrice de transition $\mathcal{P}$.

(b) La chaîne est-elle irréductible, périodique, ergodique ?

(c) Montrer que la loi de probabilité $u = (u_1, \dots, u_9)$, où chaque u_i est proportionnel au nombre de portes que comporte le compartiment i, est une loi de probabilité stationnaire pour la chaîne de Markov. Interprétation.

6.5 Les coccinelles

On considère un hexagone régulier dont les sommets sont numérotés $1, \dots, 6$, dans le sens trigonométrique. À la date 0, deux coccinelles sont placées aux sommets i et j, respectivement. À la date 1, chacune des coccinelles se déplace, indépendamment de l'autre, vers l'un des deux sommets adjacents, avec une probabilité $\frac{1}{2}$. À la date 2, l'opération se répète et ainsi de suite. On demande le

temps moyen que mettront ces coccinelles pour se rencontrer *pour la première fois*. Si $i = j$, ce temps est celui de leurs premières retrouvailles.

6.6 Soient (X_n) $(n \geq 0)$ une chaîne de Markov, homogène, à un nombre fini d'états et $\mathcal{P} = (p_{i,j})$ sa matrice de transition.

(a) On définit le processus (Y_n) $(n \geq 1)$, où $Y_n := (X_{n-1}, X_n)$. Ce processus est-il de Markov ? Si oui, déterminer sa matrice de transition.

(b) On suppose que le processus (X_n) $(n \geq 0)$ est ergodique et on désigne par $u := (u_0, u_1, \ldots, u_r)$ sa loi de probabilité stationnaire (unique). Calculer $\lim_n \mathrm{P}\{Y_n = (i,j)\}$.

6.7 On reprend les mêmes notations que dans le paragraphe 2 pour le problème de la ruine du joueur. Il y a trois classes indécomposables : $\{0\}$, $\{a\}$ (classes récurrentes, puisque 0 et a sont absorbants), $\{1, \ldots, (a-1)\}$ (classe transiente, périodique de période 2). Il existe une infinité de lois de probabilité stationnaires, portées par les deux états absorbants 0 et a.

6.8 Collectionneur de vignettes, autre calcul

Pierre achète des tablettes de chocolat d'une certaine marque. Dans chaque tablette, il trouve une vignette qu'il peut coller dans un album édité par cette marque. L'album contient m emplacements. Soit T le nombre de tablettes qu'il lui faut acheter pour terminer son album, en supposant que chaque fois la probabilité d'avoir une vignette donnée est $1/m$. Comme démontré dans l'Exercice 2.6, on a $\mathbb{E}[T] = m\, H_m$, où $H_m = 1 + \frac{1}{2} + \cdots + \frac{1}{m}$. Construire une chaîne de Markov, permettant, en faisant usage du Théorème 6.2.3, de retrouver cette évaluation.

6.9 Soit Rec (resp. Tra) l'ensemble des états récurrents (resp. transients) d'une chaîne de Markov finie et homogène et soit T le temps d'atteinte dans l'ensemble Rec.

(a) Pour tout état i, on a : $\mathbb{E}^i[T] = 1 + \sum_{j \in \mathrm{Tra}} p_{i,j}\, \mathbb{E}^j[T]$.

(b) On considère le modèle de la ruine du joueur (*cf.* § 5.2.4) pour $a = 4$ et on désigne par T le temps d'atteinte dans l'un des états de ruine 0 ou $a = 4$. Évaluer $E^i[T]$ pour $i = 1, 2, 3$.

(c) Dans le problème du jeu de tennis (*cf.* Exercice 6.1), en utilisant les mêmes hypothèses et notations, déterminer la durée moyenne de la fin du jeu à partir du moment où le tableau d'affichage montre le score $(40, 15)$.

(d) On peut aussi calculer la *fonction génératrice* de cette durée moyenne, par une analyse directe du graphe de la chaîne.

6.10 Soit $\mathcal{P} = (p_{i,j})$ $(1 \leq i,j \leq r)$ une matrice stochastique d'ordre r. On pose $\omega := \min_{i=1}^{n} p_{i,i}$, une quantité comprise entre 0 et 1. Si λ est une valeur propre de $\mathcal{P}$, alors $|\lambda - \omega| \leq 1 - \omega$, c'est-à-dire toute valeur propre de $\mathcal{P}$ est située dans le disque de centre $(\omega, 0)$ et de rayon $(1 - \omega)$.

Chapitre 7

Martingales

L'étude des martingales qui est proposée ci-dessous repose très fortement sur l'algèbre des espérances conditionnelles, dont la forme utile a été rappelée dans le paragraphe 1.13. Les *Règles* numérotées de 1 à 6, qui y ont été données, sont utilisées constamment. Comme rappelé dans le paragraphe 1.13, les formules qui comportent des espérances conditionnelles ne sont vraies que presque sûrement.

7.1 PREMIÈRES PROPRIÉTÉS

On suppose donné un processus stochastique à temps discret, c'est-à-dire une suite (X_n) $(n \geq 0)$ de variables aléatoires définies sur le même espace probabilisé.

Définition 7.1.1. On dit que ce processus est une *martingale* si, pour tout $n \geq 0$, les propriétés suivantes sont satisfaites :

(a) $\mathbb{E}[\,|X_n|\,] < +\infty$;

(b) $\mathbb{E}[X_{n+1} \,|\, X_0, X_1, \ldots, X_n] = X_n$.

Le second axiome fait apparaître l'espérance conditionnelle d'une variable aléatoire en un vecteur aléatoire. Cette définition sous-entend donc que pour tout $n \geq 0$ on sait déterminer la loi de probabilité du vecteur $(X_0, X_1, \ldots, X_n)$.

Originellement, la notion de martingale a été introduite pour décrire les caractéristiques d'un jeu *équitable*. Pour chaque n, on note X_n la fortune d'un joueur à la $n^{\text{ième}}$ partie d'un jeu donné. Le second axiome dit que, conditionnellement à la connaissance qu'a le joueur de l'évolution de sa fortune

jusqu'à l'instant n (y compris la $n^{\text{ième}}$ partie), l'espérance mathématique de sa fortune à la $(n+1)^{\text{ième}}$ partie est égale à sa fortune à la $n^{\text{ième}}$ partie.

Cette définition est un peu restrictive. On préfère se donner deux suites (X_n) et (Y_n) $(n \geq 0)$ de variables aléatoires définies sur un même espace probabilisé et prendre la définition suivante.

Définition 7.1.2. On dit que le processus (X_n) $(n \geq 0)$ est une *martingale par rapport au processus* (Y_n) $(n \geq 0)$, si, pour tout $n \geq 0$, les propriétés suivantes sont satisfaites :

(a) $\mathbb{E}[\,|X_n|\,] < +\infty$;
(b) $\mathbb{E}[X_{n+1} \,|\, Y_0, Y_1, \dots, Y_n] = X_n$.

Dans l'interprétation d'une suite de parties considérée plus haut, on peut interpréter $(Y_0, Y_1, \dots, Y_n)$ comme l'*information recueillie par le joueur jusqu'à la* $n^{\text{ième}}$ *partie* (cette partie incluse). Or cette information peut comporter des données autres que la suite des fortunes $X_0, X_1, \dots, X_n$, comme, par exemple, les résultats des parties où le joueur n'a pas joué.

D'après la Proposition 1.13.1, on voit que l'on peut remplacer l'axiome (b) de la Définition 7.1.2 par :

(b′) pour tout entier $n \geq 0$ et tout évènement A appartenant à la tribu $\mathfrak{T}(Y_0, Y_1, \dots, Y_n)$ engendrée par le vecteur $(Y_0, Y_1, \dots, Y_n)$, on a la relation :

$$\mathbb{E}[I_A X_n] = \mathbb{E}[I_A X_{n+1}].$$

C'est, en effet, traditionnellement sous cette forme que l'axiome (b) est énoncé.

Propriété 7.1.1. *Si (X_n) est une martingale par rapport à (Y_n), alors X_n est une fonction mesurable de $(Y_0, \dots, Y_n)$. (On dit que la suite (X_n) est adaptée à la suite (Y_n).)*

Démonstration. Cette propriété découle immédiatement de la définition de l'espérance conditionnelle. Se reportant au diagramme de définition, on a bien $\mathbb{E}[X_{n+1} \,|\, Y_0, \dots, Y_n] = e(Y_0, \dots, Y_n)$, où e est une fonction mesurable de $\mathbb{R}^{n+1}$ dans $\mathbb{R}$. □

Propriété 7.1.2. *Si (X_n) est une martingale par rapport à (Y_n), alors, pour tout $n \geq 0$, on a : $\mathbb{E}[X_n] = \mathbb{E}[X_0]$. (Une martingale est à espérance mathématique constante.)*

Démonstration. De l'axiome (b) de la définition d'une martingale on déduit : $\mathbb{E}[X_n] = \mathbb{E}[\mathbb{E}[X_{n+1} \,|\, Y_0, Y_1, \dots, Y_n]\,]$. Or d'après la Règle 5, en prenant pour g l'application identique, on a $\mathbb{E}[\mathbb{E}[X_{n+1} \,|\, Y_0, Y_1, \dots, Y_n]\,] = \mathbb{E}[X_{n+1}]$. D'où $\mathbb{E}[X_{n+1}] = \mathbb{E}[X_n]$ et donc $\mathbb{E}[X_n] = \mathbb{E}[X_0]$ par récurrence. □

Propriété 7.1.3. *Si (X_n) est une martingale par rapport à (Y_n), alors, pour tout $n \geq 0$, on a : $\mathbb{E}[X_n \,|\, Y_0, \dots, Y_n] = X_n$.*

Démonstration. D'après la Propriété 7.1.1, on a $X_n = e(Y_0, \dots, Y_n)$. On applique alors la Règle 2 avec $h = e$. □

Propriété 7.1.4. *Si* (X_n) *est une martingale par rapport à* (Y_n)*, alors, pour tout* $n \geq 0$ *et tout* $k \geq 0$ *on a :* $\mathbb{E}[X_{n+k} \,|\, Y_0, \ldots, Y_n] = X_n$.

Démonstration. La propriété est déjà vraie pour $k = 0, 1$. Procédons par récurrence par rapport à $k \geq 1$. Dans l'identité de la Règle 6, prenons pour g l'application identique, puis faisons les substitutions $X \leftarrow X_{n+k+1}$, $Z \leftarrow (Y_0, \ldots, Y_n)$ et $Y \leftarrow (Y_{n+1}, \ldots, Y_{n+k})$. On obtient :

$$\begin{aligned}\mathbb{E}[X_{n+k+1} \,|\, (Y_0, \ldots, Y_n)] &= \mathbb{E}[\mathbb{E}[X_{n+k+1} \,|\, Y_0, \ldots, Y_{n+k}] \,|\, Y_0, \ldots, Y_n] \\ &= \mathbb{E}[X_{n+k} \,|\, Y_0, \ldots, Y_n] = X_n\end{aligned}$$

d'abord puisque (X_n) est une martingale par rapport à (Y_n), ensuite en appliquant l'hypothèse de récurrence. □

Donnons également une autre définition d'une martingale.

Définition 7.1.3. Soient (X_n) et (Y_n) $(n \geq 0)$ deux processus définis sur un même espace probabilisé. Le processus (X_n) est une *martingale par rapport au processus* (Y_n), si pour tout $n \geq 0$, les propriétés suivantes sont satisfaites :

(1) $\mathbb{E}[\,|X_n|\,] < +\infty$;
(2) X_n est une fonction mesurable de $Y_0, \ldots, Y_n$;
(3) $\mathbb{E}[(X_{n+1} - X_n) \,|\, Y_0, \ldots, Y_n] = 0$.

Les Définitions 7.1.2 et 7.1.3 sont bien équivalentes, car la Définition 7.1.2 implique la Définition 7.1.3, en vertu des Propriétés 7.1.1 et 7.1.3. D'autre part, la Définition 7.1.3 implique la Définition 7.1.2, en vertu de la Règle 2. Notons qu'à cause de la Propriété 7.1.4, on peut remplacer la condition (3) dans la Définition 7.1.3 par

(3′) pour tout $n, k \geq 0$, on a : $\mathbb{E}[(X_{n+k} - X_n) \,|\, Y_0, \ldots, Y_n] = 0$.

Les suites de sommes partielles de variables aléatoires indépendantes et centrées constituent les exemples les plus courants de martingales, comme décrit ci-après.

Proposition 7.1.5. *Soient* $Y_0 = 0$ *la variable aléatoire identiquement nulle et* (Y_n) $(n \geq 1)$ *une suite de variables aléatoires mutuellement indépendantes, centrées et telles que pour tout* $n \geq 0$ *on ait* $\mathbb{E}[\,|Y_n|\,] < +\infty$*. Pour tout* $n \geq 0$*, on pose* $X_n := Y_0 + Y_1 + \cdots + Y_n$*. Alors la suite* (X_n) $(n \geq 0)$ *est une martingale par rapport à* (Y_n) $(n \geq 0)$.

Démonstration. Tout d'abord, $\mathbb{E}[\,|X_n|\,] \leq \mathbb{E}[\,|Y_0|\,] + \mathbb{E}[\,|Y_1|\,] + \cdots + \mathbb{E}[\,|Y_n|\,]$, une somme qui est finie. Ensuite,

$$\mathbb{E}[X_{n+1} \,|\, Y_0, \ldots, Y_n] = \mathbb{E}[X_n \,|\, Y_0, \ldots, Y_n] + \mathbb{E}[Y_{n+1} \,|\, Y_0, \ldots, Y_n].$$

En prenant $h(y_0, \ldots, y_n) = y_0 + \cdots + y_n$ dans la Règle 2, on obtient : $\mathbb{E}[X_n \,|\, Y_0, \ldots, Y_n] = Y_0 + \cdots + Y_n = X_n$. D'autre part, comme les variables Y_{n+1} et $(Y_0, \ldots, Y_n)$ sont indépendantes, la Règle 1 donne : $\mathbb{E}[Y_{n+1} \,|\, Y_0, \ldots, Y_n] = \mathbb{E}[Y_{n+1}] = 0$. D'où $\mathbb{E}[X_{n+1} \,|\, Y_0, \ldots, Y_n] = X_n$. □

7.2 LA MARTINGALE DE DOOB

On part d'une variable aléatoire X de référence, du premier ordre, c'est-à-dire telle que $\mathbb{E}[|X|] < +\infty$. On se donne, par ailleurs, une suite (Y_n) $(n \geq 0)$ de variables aléatoires, définies sur le même espace probabilisé que X. On se propose d'étudier le processus (X_n) $(n \geq 0)$ avec $X_n := \mathbb{E}[X \mid Y_0, \dots, Y_n]$.

Proposition 7.2.1. *La suite (X_n) $(n \geq 0)$ est une martingale par rapport à la suite (Y_n) $(n \geq 0)$. On l'appelle martingale de Doob. Cette martingale peut être considérée comme modèle de phénomènes d'apprentissage.*

Démonstration. D'abord,

$$\mathbb{E}[|X_n|] = \mathbb{E}[|\mathbb{E}[X \mid Y_0, \dots, Y_n]|] \leq \mathbb{E}[\mathbb{E}[|X| \mid Y_0, \dots, Y_n]] = \mathbb{E}[|X|] < +\infty,$$

d'après la Règle 5.
Ensuite,

$$\begin{aligned} \mathbb{E}[X_{n+1} \mid Y_0, \dots, Y_n] &= \mathbb{E}[\mathbb{E}[X \mid Y_0, \dots, Y_n, Y_{n+1}] \mid Y_0, \dots, Y_n] \\ &= \mathbb{E}[X \mid Y_0, \dots, Y_n] \qquad \text{[d'après la Règle 6]} \\ &= X_n. \quad \square \end{aligned}$$

Dans le cas particulier où $X = I_A$, on a $X_n = \mathrm{P}(A \mid Y_0, \dots, Y_n)$. La suite (X_n) $(n \geq 0)$ est donc une martingale de Doob.

Nous nous proposons de voir comment toute martingale peut s'exprimer comme une martingale de Doob. Dans ce but, nous supposons donné une variable aléatoire Z uniformément répartie sur l'intervalle $[0, 1]$. Pour chaque entier $n \geq 0$ on définit ensuite la variable Y_n comme étant la $n^{\text{ième}}$ approximation de Z dans le développement binaire de Z. En d'autres termes, pour $k = 0, 1, \dots, 2^n - 1$, on pose :

$$\{Y_n = k/2^n\} \quad \text{si et seulement si} \quad \{k/2^n \leq Z < (k+1)/2^n\}.$$

Enfin, si f est une fonction bornée sur $[0, 1]$, on forme le quotient des différences :

$$X_n := \frac{f \circ (Y_n + 1/2^n) - f \circ Y_n}{1/2^n}. \tag{7.2.1}$$

Proposition 7.2.2. *La suite (X_n) $(n \geq 0)$ est une martingale par rapport à (Y_n) $(n \geq 0)$.*

Démonstration. D'abord, on a $\mathbb{E}[|X_n|] < +\infty$, puisque f est bornée. Ensuite, par définition des Y_n, conditionnellement à $(Y_0, \dots, Y_n)$, la variable aléatoire Z prend ses valeurs dans $[Y_n, Y_n + 1/2^n[$. Par ailleurs, pour tout sous-ensemble J de $[Y_n, Y_n + 1/2^n[$, on a

$$\begin{aligned} \mathrm{P}\{Z \in J \mid Z \in [Y_n, Y_n + 1/2^n[\} &= \frac{\mathrm{P}\{Z \in J\}}{\mathrm{P}\{Z \in [Y_n, Y_n + 1/2^n[\}} \\ &= 2^n\, \mathrm{P}\{Z \in J\} = 2^n\, \ell(J), \end{aligned}$$

où $\ell(J)$ désigne la longueur de J. Ceci montre que, conditionnellement à $(Y_0, \ldots, Y_n)$, la variable aléatoire Z est uniformément répartie sur $[Y_n, Y_n + 1/2^n[$.

D'autre part, conditionnellement à $(Y_0, \ldots, Y_n)$, la variable aléatoire Y_{n+1} prend l'une des deux valeurs Y_n, $Y_n + 1/2^{n+1}$. Elle prend la valeur Y_n, avec la probabilité $2^n/2^{n+1} = 1/2$, si $Z \in [Y_n, Y_n + 1/2^{n+1}[$ et la valeur $Y_n + 1/2^{n+1}$ avec la probabilité $2^n/2^{n+1} = 1/2$, si $Z \in [Y_n + 1/2^{n+1}, Y_n + 1/2^n[$. Ainsi, conditionnellement à $(Y_0, Y_1, \ldots, Y_n)$, la variable aléatoire Y_{n+1} prend l'une ou l'autre des valeurs Y_n, $Y_n + 1/2^{n+1}$ avec la probabilité $\frac{1}{2}$.
Il en résulte

$$\begin{aligned}
\mathbb{E}[X_{n+1} \,|\, Y_0, \ldots, Y_n] &= 2^{n+1}\, \mathbb{E}[f \circ \Big(Y_{n+1} + \frac{1}{2^{n+1}}\Big) - f \circ Y_{n+1} \,|\, Y_0, \ldots, Y_n] \\
&= 2^{n+1} \Big[\frac{1}{2}\Big(f \circ \Big(Y_n + \frac{1}{2^{n+1}}\Big) - f \circ Y_n\Big) \\
&\qquad + \frac{1}{2}\Big(f \circ \Big(Y_n + \frac{1}{2^n}\Big) - f \circ \Big(Y_n + \frac{1}{2^{n+1}}\Big)\Big)\Big] \\
&= 2^n \Big(f \circ \Big(Y_n + \frac{1}{2^n}\Big) - f \circ Y_n\Big) = X_n. \quad \square
\end{aligned}$$

L'expression de X_n, donnée en (7.2.1), peut être interprétée comme une dérivée discrète de f en Z. Avec des hypothèses de régularité convenables, nous nous proposons de démontrer maintenant que la suite (X_n) de ces dérivées discrètes converge presque sûrement vers une limite, notée X_∞, appelée *dérivée de Radon-Nikodym de f en Z*.

Proposition 7.2.3. *Supposons que f soit continue au sens de Lipschitz, à savoir qu'il existe une constante C telle que, pour tout x et tout y pris dans l'intervalle $[0, 1]$, on ait $|f(x) - f(y)| \leq C|x - y|$. Alors la suite (X_n) $n \geq 0)$ converge presque sûrement vers une variable aléatoire limite X_∞, appelée la dérivée de Radon-Nikodym de f en Z.*

Démonstration. D'abord, la suite (X_n) $(n \geq 0)$ est une martingale par rapport à (Y_n) $(n \geq 0)$ d'après la Proposition 7.2.2. Pour établir la proposition, on utilise le théorème de convergence pour les martingales (qu'on établit dans les paragraphes suivants) qui dit que si (X_n) est une martingale satisfaisant $\sup_{n \geq 0} \mathbb{E}[\,|X_n|\,] < +\infty$, alors (X_n) converge presque sûrement vers une variable aléatoire X_∞ intégrable. L'inégalité précédente est évidemment satisfaite, puisque, d'après la condition de Lipschitz, on a, pour tout $n \geq 0$,

$$|X_n| = \frac{1}{1/2^n} |f \circ (Y_n + 1/2^n) - f \circ Y_n| < C < +\infty,$$

et donc $\sup_{n \geq 0} \mathbb{E}[\,|X_n|\,] < +\infty$. $\square$

La variable X_∞ permet d'exprimer la martingale (X_n) comme une martingale de Doob, comme montré dans la proposition suivante.

Proposition 7.2.4. *On conserve la définition de X_n donnée en* (7.2.1) *et les hypothèses de la Proposition* 7.2.3. *Soit, de plus, X_∞ la dérivée de Radon-Nikodym de f en Z. Alors, pour tout $n \geq 0$, on a :*

$$X_n = \mathbb{E}[X_\infty \,|\, Y_0, \ldots, Y_n]. \tag{7.2.2}$$

Démonstration. On a vu, en vertu de la condition de Lipschitz, que pour tout $n \geq 0$ on a : $|X_n| \leq C < +\infty$. Il en résulte, et nous l'admettrons, que la suite (X_n) est une *martingale de Doob*. Comme on a, par ailleurs, $X_n \xrightarrow{p.s.} X_\infty$, cette martingale (X_n) admet la représentation (7.2.2). □

7.3 LA MARTINGALE DE WALD

Cette martingale intervient dans de nombreuses applications et mérite un traitement spécial.

Proposition 7.3.1. *Soient $Y_0 := 0$ et (Y_n) $(n \geq 1)$ une suite de variables aléatoires, indépendantes, identiquement distribuées, admettant une fonction génératrice $g(u) = \mathbb{E}[e^{uY}]$ pour au moins un réel $u \neq 0$. Posons $X_0 := 1$ et pour $n \geq 1$ posons $X_n := g(u)^{-n} e^{u(Y_1+\cdots+Y_n)}$. Alors (X_n) $(n \geq 0)$ est une martingale par rapport à (Y_n) $(n \geq 0)$, appelée martingale de Wald.*

Démonstration. D'abord, pour tout $n \geq 1$, on a : $\mathbb{E}[\,|X_n|\,] = \mathbb{E}[X_n] = 1 < +\infty$. Notons qu'on a ensuite $X_{n+1} = X_n \, e^{uY_{n+1}}/g(u)$, d'où

$$\begin{aligned}
\mathbb{E}[X_{n+1} \,|\, Y_0, \ldots, Y_n] &= \mathbb{E}[X_n \frac{e^{uY_{n+1}}}{g(u)} \,|\, Y_0, \ldots, Y_n] \\
&= X_n \mathbb{E}\Big[\frac{e^{uY_{n+1}}}{g(u)} \,|\, Y_0, \ldots, Y_n\Big] \quad \text{[en vertu de la Règle 4]} \\
&= X_n \mathbb{E}\Big[\frac{e^{uY_{n+1}}}{g(u)}\Big] = X_n. \quad \square
\end{aligned}$$

[en vertu de l'indépendance des Y_n]

Exemple. Prenons $\mathcal{L}(Y) = N(0, \sigma)$, d'où $g(u) = \mathbb{E}[e^{uY}] = e^{u^2\sigma^2/2}$. On voit que $g(u)$ est définie pour tout $u \in \mathbb{R}$. Il en résulte que $X_n = e^{u(Y_1+\cdots+Y_n) - nu^2\sigma^2/2}$ est une martingale pour tout $u \in \mathbb{R}$.

7.4 SURMARTINGALES ET SOUS-MARTINGALES

Rappelons que pour tout nombre réel x, on pose

$$x^+ := \sup(x, 0) \quad \text{et} \quad x^- := -\inf(x, 0).$$

Définition 7.4.1. Soient (X_n), (Y_n) $(n \geq 0)$ deux processus définis sur un même espace probabilisé. On dit que (X_n) est une *surmartingale* par rapport à (Y_n), si pour tout $n \geq 0$,

(a) $\mathbb{E}[X_n^-] < +\infty$;
(b) $\mathbb{E}[X_{n+1} \,|\, Y_0, Y_1, \ldots, Y_n] \le X_n$;
(c) X_n est une fonction mesurable de $Y_0, \ldots, Y_n$.

On dit que (X_n) est une *sous-martingale* par rapport à (Y_n), si pour tout $n \ge 0$,

(a) $\mathbb{E}[X_n^+] < +\infty$;
(b) $\mathbb{E}[X_{n+1} \,|\, Y_0, Y_1, \ldots, Y_n] \ge X_n$;
(c) X_n est une fonction mesurable de $Y_0, \ldots, Y_n$.

Remarque 1. On note ici que la condition (c) était implicite dans la Définition 7.1.1 des martingales et résultait de l'axiome (b), ce qui n'est pas le cas pour les surmartingales et les sous-martingales.

Remarque 2. Il est évident que (X_n) est une surmartingale si et seulement si $(-X_n)$ est une sous-martingale par rapport à (Y_n). Par ailleurs, (X_n) est une martingale si et seulement si (X_n) est à la fois une sur-martingale et une sous-martingale par rapport à (Y_n).

Donnons des méthodes de construction de sous-martingales à partir d'une martingale ou d'autres sous-martingales.

Lemme 7.4.1. *Supposons données :*

(a) *une martingale (X_n) par rapport à (Y_n) ;*

(b) *une fonction convexe φ telle que, pour tout $n \ge 0$, l'espérance mathématique $\mathbb{E}[(\varphi \circ X_n)^+]$ est finie.*

Alors $(\varphi \circ X_n)$ est une sous-martingale par rapport à (Y_n).

Démonstration. On rappelle l'*inégalité de Jensen* qui stipule que si φ est une fonction convexe et si X est une variable aléatoire d'espérance mathématique finie, on a : $\varphi(\mathbb{E}[X \,|\, Y]) \le \mathbb{E}[\varphi \circ X \,|\, Y]$ (voir Exercice 7.8).

Il s'agit de vérifier l'axiome (b) des sous-martingales. D'après l'inégalité de Jensen, on a $\mathbb{E}[\varphi \circ X_{n+1} \,|\, Y_0, \ldots, Y_n] \ge \varphi(\mathbb{E}[X_{n+1} \,|\, Y_0, \ldots, Y_n]) = \varphi(X_n)$, puisque (X_n) est une martingale par rapport à (Y_n). □

Corollaire. *Soit (X_n) une martingale par rapport à (Y_n). Alors*

(a) *$(|X_n|)$ est une sous-martingale par rapport à (Y_n) ;*

(b) *si pour tout $n \ge 0$ on a $\mathbb{E}[X_n^2] < +\infty$, alors (X_n^2) est une sous-martingale par rapport à (Y_n).*

Démonstration. Pour (a), on prend la fonction convexe $x \mapsto |x|$ et pour (b) la fonction $x \mapsto x^2$. Dans ce dernier cas, l'existence du second moment des X_n est indispensable pour que l'axiome (a) des sous-martingales soit satisfait. □

Lemme 7.4.2. *Soient (X_n) une sous-martingale par rapport à (Y_n) et φ une fonction convexe et croissante telle que pour tout $n \ge 0$ on ait $\mathbb{E}[\,|\varphi \circ X_n|\,] < +\infty$. Alors $(\varphi \circ X_n)$ est une sous-martingale par rapport à (Y_n).*

Démonstration. Il s'agit de vérifier l'axiome (b) de la définition des sous-martingales. Or, en appliquant l'inégalité de Jensen, puis en utilisant le fait que (X_n) est une sous-martingale et que φ est croissante, on obtient :

$$\mathbb{E}[\varphi \circ X_{n+1} \mid Y_0, \ldots, Y_n] \geq \varphi(\mathbb{E}[X_{n+1} \mid Y_0, \ldots, Y_n]) \geq \varphi \circ X_n. \quad \square$$

Corollaire. *Si (X_n) $(n \geq 0)$ est une sous-martingale, il en est de même de (X_n^+) $(n \geq 0)$.*

Démonstration. Plus généralement, prenons $\varphi(x) := x$ ou c suivant que l'on a $x \geq c$ ou non. Alors φ est croissante et convexe. De plus, $\mathbb{E}[\,|\varphi \circ X_n|\,] \leq \mathbb{E}[\,|X_n|\,] < +\infty$. Le Lemme 7.4.2 s'applique et $(\varphi \circ X_n)$ est une sous-martingale par rapport à (Y_n). L'énoncé du Corollaire correspond au cas particulier $c = 0$. $\square$

Nous donnons, maintenant, pour les sous-martingales et les surmartingales, les propriétés analogues à celles énoncées dans les Propriétés 7.1.2-7.1.4 pour les martingales.

Propriété 7.4.3. *Soit (X_n) $(n \geq 0)$ une martingale (resp. une sous-martingale, resp. une surmartingale) par rapport à (Y_n) $(n \geq 0)$. Alors la suite numérique de terme général $\mathbb{E}[X_n]$ est constante (resp. croissante, resp. décroissante).*

Démonstration. Si (X_n) est une martingale (resp. une sous-martingale, resp. une surmartingale), on a : $\mathbb{E}[X_{n+1} - X_n \mid Y_0, \ldots, Y_n] = 0$ (resp. ≥ 0, resp. ≤ 0). En prenant l'espérance mathématique, on obtient, d'après (1.2), les relations $\mathbb{E}[X_{n+1} - X_n] = 0$ (resp. ≥ 0, resp. ≤ 0), ou encore $\mathbb{E}[X_{n+1}] = \mathbb{E}[X_n] = \mathbb{E}[X_0]$ (resp. $\mathbb{E}[X_n] \leq \mathbb{E}[X_{n+1}]$, resp. $\mathbb{E}[X_n] \geq \mathbb{E}[X_{n+1}]$). $\square$

Propriété 7.4.4. *Soit (X_n) $(n \geq 0)$ une martingale (resp. une sous-martingale, resp. une surmartingale) par rapport à (Y_n) $(n \geq 0)$. Alors, pour tout $k \geq 0$ on a : $\mathbb{E}[X_{n+k} \mid (Y_0, \ldots, Y_n)] = X_n$ (resp. $\geq X_n$, resp. $\leq X_n$).*

Démonstration. On procède par récurrence sur k, la propriété étant déjà vraie pour $k = 0, 1$. En appliquant la Règle 6, prenons $g \circ X := X_{n+k+1}$, $Z := (Y_0, \ldots, Y_n)$, $Y := (Y_{n+1}, \ldots, Y_{n+k})$. Alors

$$\mathbb{E}[\mathbb{E}[X_{n+k+1} \mid (Y_0, \ldots, Y_{n+k})] \mid (Y_0, \ldots, Y_n)] = \mathbb{E}[X_{n+k+1} \mid (Y_0, \ldots, Y_n)].$$

Or, si (X_n) est une martingale (resp. une sous-martingale, resp. une surmartingale), on a : $\mathbb{E}[X_{n+k+1} \mid Y_0, \ldots, Y_{n+k}] = X_{n+k}$ (resp. $\geq X_{n+k}$, resp. $\leq X_{n+k}$). D'où $\mathbb{E}[X_{n+k+1} \mid (Y_0, \ldots, Y_n)] = \mathbb{E}[X_{n+k} \mid (Y_0, \ldots, Y_n)]$ (resp. $\mathbb{E}[X_{n+k+1} \mid (Y_0, \ldots, Y_n)] \geq \mathbb{E}[X_{n+k} \mid (Y_0, \ldots, Y_n)]$, resp. $\mathbb{E}[X_{n+k+1} \mid (Y_0, \ldots, Y_n)] \leq \mathbb{E}[X_{n+k} \mid (Y_0, \ldots, Y_n)]$). D'où, enfin, $\mathbb{E}[X_{n+k+1} \mid (Y_0, \ldots, Y_n)] = X_n$ (resp. $\geq X_n$, resp. $\leq X_n$), par l'hypothèse de récurrence. $\square$

Proposition 7.4.5. *Soit (X_n) $(n \geq 0)$ une martingale (resp. une sous-martingale, resp. une surmartingale) par rapport à (Y_n) $(n \geq 0)$. Soit A_n un évènement de la tribu engendrée par $(Y_0, \ldots, Y_n)$. (On peut encore dire un*

évènement dont l'indicatrice I_{A_n} est une fonction mesurable de $(Y_0,\ldots,Y_n)$.) Alors, pour tout $k \geq 0$ on a : $\mathbb{E}[I_{A_n}X_{n+k}] = \mathbb{E}[I_{A_n}X_n]$ (resp. $\mathbb{E}[I_{A_n}X_{n+k}] \geq \mathbb{E}[I_{A_n}X_n]$, resp. $\mathbb{E}[I_{A_n}X_{n+k}] \leq \mathbb{E}[I_{A_n}X_n]$).

Démonstration. D'après la Proposition 7.4.4, on a :

$$\begin{aligned}\mathbb{E}[I_{A_n}X_{n+k}] &= \mathbb{E}[\mathbb{E}[I_{A_n}X_{n+k} \mid Y_0,\ldots,Y_n]] = \mathbb{E}[I_{A_n}\,\mathbb{E}[X_{n+k} \mid Y_0,\ldots,Y_n]] \\ &= \mathbb{E}[I_{A_n}\,X_n] \text{ (resp. } \geq \mathbb{E}[I_{A_n}\,X_n]\text{, resp. } \leq \mathbb{E}[I_{A_n}\,X_n]). \quad \square\end{aligned}$$

7.5 COMPLÉMENTS ET EXERCICES

7.1 Soit Y_0 la variable identiquement nulle et (Y_n) $(n \geq 1)$ une suite de variables aléatoires, *mutuellement indépendantes, identiquement distribuées, centrées et ayant chacune le moment du second ordre* σ^2. Pour chaque $n \geq 0$ on pose :

$$X_n := \Big(\sum_{k=0}^{n} Y_k\Big)^2 - \mathbb{E}[\Big(\sum_{k=0}^{n} Y_k\Big)^2] = \Big(\sum_{k=0}^{n} Y_k\Big)^2 - n\,\sigma^2.$$

Alors (X_n) $(n \geq 0)$ est une martingale par rapport à (Y_n) $(n \geq 0)$.

7.2 Martingale associée à une chaîne de Markov

Soit (Y_n) $(n \geq 0)$ une chaîne de Markov, dont l'ensemble des états est $E = \mathbb{N}^* = \mathbb{N} \setminus \{0\}$ et la matrice de transition est notée $\mathcal{P} = (p_{ij})$ $(i,j \geq 0)$. Soit f une application de $\mathbb{N}^*$ dans $\mathbb{R}+$, telle que pour tout $n \geq 0$ on ait $\mathbb{E}[|f \circ Y_n|] < \infty$, non identiquement nulle, vérifiant l'identité $\mathcal{P}f = f$. Autrement dit, f est un vecteur propre à droite de $\mathcal{P}$, associée à la valeur propre 1, ou encore satisfait le système d'équations : $\sum_j p_{ij}\,f(j) = f(i)$ $(i \geq 0)$.

Pour chaque $n \geq 0$, on pose : $X_n := f \circ Y_n$. Alors (X_n) $(n \geq 0)$ est une martingale par rapport à (Y_n) $(n \geq 0)$.

7.3 On reprend les mêmes hypothèses que dans l'énoncé du précédent exercice, avec la seule différence que l'on suppose que f vérifie l'identité $\mathcal{P}f = \lambda f$ pour un certain nombre $\lambda \neq 0$; de plus, pour chaque $n \geq 0$ on pose cette fois : $X_n = \lambda^{-n} f \circ Y_n$. Alors (X_n) $(n \geq 0)$ est une martingale par rapport à (Y_n) $(n \geq 0)$.

7.4 La martingale de Wald ; cas particulier

Soit Y une variable aléatoire prenant ses valeurs dans $\mathbb{Z}$. Pour tout $k \in \mathbb{Z}$, on pose $q_k := \mathrm{P}\{Y = k\}$. On suppose que la fonction génératrice des moments $g(u) = \mathbb{E}[e^{uY}]$ existe pour au moins une valeur $u \neq 0$. Soient $Y_0 := 0$ et

(Y_n) $(n \geq 1)$ une suite de variables aléatoires, indépendantes, identiquement distribuées, ayant chacune même loi de probabilité que Y.

(a) On pose $S_0 := 0$ et $S_n := Y_1 + \cdots + Y_n$ pour $n \geq 1$. Alors (S_n) $(n \geq 0)$ est une chaîne de Markov, d'ensemble des états $\mathbb{Z}$ et de matrice de transition $\mathcal{P} = (q_{j-i})$ $(i, j \in \mathbb{Z})$.

(b) On a $\mathcal{P}f = g(u)\,f$ pour la fonction $f : j \mapsto e^{uj}$.

(c) Enfin, on pose $X_0 := 1$ et pour $n \geq 1$, on définit

$$X_n := g(u)^{-n} \exp(u(Y_1 + \cdots + Y_n)).$$

Alors (X_n) $(n \geq 0)$ est une martingale par rapport à (Y_n) $(n \geq 0)$.

7.5 Rapports de vraisemblance

Soient (Y_n) $(n \geq 0)$ une suite de variables aléatoires, indépendantes, identiquement distribuées, admettant chacune la densité de probabilité f_0, ne prenant pas la valeur nulle. Soit f_1 une seconde densité de probabilité. On considère la suite des rapports de vraisemblance (X_n) $(n \geq 0)$, où, pour chaque $n \geq 0$, on pose :

$$X_n := \frac{f_1 \circ Y_0 \cdot f_1 \circ Y_1 \cdots f_1 \circ Y_n}{f_0 \circ Y_0 \cdot f_0 \circ Y_1 \cdots f_0 \circ Y_n}.$$

Si les variables aléatoires Y_i admettent toutes $f_0 > 0$ comme densité de probabilité, alors (X_n) $(n \geq 0)$ est une martingale par rapport à (Y_n) $(n \geq 0)$.

7.6 Soient $Y_0 = c$ et (Y_n) $(n \geq 1)$ une suite de variables aléatoires, indépendantes, telles que pour tout $n \geq 1$ on ait : $\mathbb{E}[\,|Y_n|\,] < +\infty$ et $\mathbb{E}[Y_n] = 1$. On pose $X_0 := c$ et $X_n := cY_1 \cdots Y_n$ $(n \geq 1)$. Alors (X_n) $(n \geq 0)$ est une martingale par rapport à (Y_n) $(n \geq 0)$.

7.7 La martingale de Wald ; autre démonstration

On prend les mêmes hypothèses que dans l'Exercice 7.4. En utilisant les techniques de l'Exercice 7.6, on démontre directement que (X_n) $(n \geq 0)$ est une martingale par rapport à (Y_n) $(n \geq 0)$.

7.8 L'inégalité de Jensen

Soient φ une fonction convexe de $\mathbb{R}$ dans $\mathbb{R}$ et X une variable aléatoire telle que X et $\varphi \circ X$ aient des espérances mathématiques finies. Alors $\varphi(\mathbb{E}[X]) \leq \mathbb{E}[\varphi \circ X]$. L'inégalité est aussi vraie pour les espérances conditionnelles : $\varphi(\mathbb{E}[X \mid Y]) \leq \mathbb{E}[\varphi \circ X \mid Y]$.

[Utiliser le fait que le graphe de φ est au-dessus de toute droite d'appui, autrement dit, pour tout ξ il existe un nombre $\lambda = \lambda(\xi)$ tel que la droite d'appui passant par $(\xi, \varphi(\xi))$ et d'équation $y - \varphi(\xi) = \lambda(x - \xi)$ soit au-dessous du graphe de φ, soit $\varphi(x) \geq \lambda(x - \xi) + \varphi(\xi)$.]

7.9 Soit (Y_n) $(n \geq 1)$ une suite de variables aléatoires indépendantes et identiquement distribuées et telles que pour tout $n \geq 1$ on ait $\mathbb{E}[\,|Y_n|\,] < +\infty$. On note μ la valeur commune de l'espérance mathématique et pour $n \geq 1$ on pose $X_n := \sum_{k=1}^{n} Y_k$. Alors la suite (X_n) $(n \geq 1)$ est une sous-martingale, une surmartingale, une martingale par rapport à (Y_n) $(n \geq 1)$, suivant que μ est ≥ 0, ≤ 0, $= 0$.

7.10 Soit (Y_n) $(n \geq 1)$ une suite de variables aléatoires indépendantes, centrées, admettant chacune un moment du second ordre. On pose $S_n := \sum_{k=1}^{n} Y_k$ et $X_n := (S_n)^2$. Alors la suite (X_n) $(n \geq 1)$ est une sous-martingale par rapport à (Y_n).

7.11 Soit (X_n) $(n \geq 1)$ une suite de variables aléatoires, indépendantes, identiquement distribuées, du premier ordre. On pose $S_n := \sum_{k=1}^{n} X_k$, $Y_n := S_n/n$ $(n \geq 1)$. Alors la suite $\dots, Y_n, Y_{n-1}, \dots, Y_1$ est une martingale par rapport à elle-même.

Chapitre 8

Théorèmes d'arrêt

On se propose, dans ce chapitre, de savoir comment lorsqu'on remplace le temps linéaire, matérialisé par l'indice n, dans une martingale (X_n), par un *temps d'arrêt* T adapté aux X_n, la variable aléatoire transformée (X_T) a encore comme espérance mathématique la valeur commune des espérances des X_n. Le résultat le plus simple à établir et qui pourtant est à la base de tous les résultats ultérieurs est obtenu lorsque le temps d'arrêt T est *borné* (*cf.* Théorème 8.2.1). Il constitue le premier énoncé du *théorème d'arrêt.* Nous donnons encore trois autres variantes de ce théorème.

8.1 PROCESSUS ARRÊTÉ À UN INSTANT

Considérons un processus stochastique (Y_n) $(n \geq 0)$ défini sur un espace probabilisé $(\Omega, \mathfrak{A}, \mathrm{P})$. Rappelons que l'on note $\mathfrak{A}_n$ la tribu engendrée par le vecteur $(Y_0, Y_1, \ldots, Y_n)$ (*cf.* § 2.1). Rappelons également (*cf.* § 2.2) qu'une variable aléatoire T, définie sur Ω, à valeurs dans $\mathbb{N} \cup \{+\infty\}$, est un *temps d'arrêt* du processus (Y_n), si, pour tout entier $n \geq 0$, l'évènement $\{T = n\}$ appartient à la tribu $\mathfrak{A}_n$.

On se donne deux processus $X = (X_n)$ et $Y = (Y_n)$ $(n \geq 0)$ définis sur un même espace probabilisé $(\Omega, \mathfrak{A}, \mathrm{P})$. On suppose que X est *adapté à* Y, ce qui veut dire que, pour tout $n \geq 0$, la variable aléatoire X_n est une fonction mesurable du vecteur $(Y_0, Y_1, \ldots, Y_n)$, ou encore (*cf.* Proposition 1.12.1) que X_n est une variable aléatoire par rapport au couple $(\Omega, \mathfrak{A}_n)$. On se donne, enfin, un temps d'arrêt T du processus Y.

On définit un nouveau processus, noté $X^{|T} = ((X^{|T})_n)$ $(n \geq 0)$, en posant, pour tout $n \geq 0$,

$$(X^{|T})_n(\omega) := X_{T\wedge n}(\omega) = \begin{cases} X_n(\omega), & \text{si } n < T(\omega)\,; \\ X_{T(\omega)}(\omega), & \text{si } n \geq T(\omega). \end{cases} \tag{8.1.1}$$

On peut encore écrire :

$$X_{T\wedge n} = X_0 I_{\{T=0\}} + X_1 I_{\{T=1\}} + \cdots + X_n I_{\{T=n\}} + X_n I_{\{T>n\}}. \tag{8.1.2}$$

Comme chacun des évènements $\{T=0\}$, $\{T=1\}$, ... , $\{T=n\}$, $\{T>n\}$ appartient à $\mathfrak{A}_n$ (*cf.* Proposition 2.2.1), leurs indicatrices sont des fonctions mesurables de $(Y_0, Y_1, \ldots, Y_n)$. Comme, par hypothèse, X_0, X_1, ... , X_n sont aussi de telles fonctions mesurables, le processus $X^{|T}$ est aussi adapté à Y. On l'appelle le *processus obtenu en arrêtant* X *à l'instant (aléatoire)* T.

Donnons une autre expression pour $X_{T\wedge n}$ que celle donnée en (8.1.2), qui est mieux adaptée pour les calculs ultérieurs.

Lemme 8.1.1. *On a la représentation :*

$$X_{T\wedge n} = X_0 + \sum_{k=0}^{n-1} I_{\{k<T\}}(X_{k+1} - X_k). \tag{8.1.3}$$

L'intérêt de cette représentation tient au fait que $I_{\{k<T\}}$ *est une fonction mesurable de* $(Y_0, \ldots, Y_k)$.

Démonstration. Tout d'abord, $X_{T\wedge n} = X_0 + \sum\limits_{0\leq k<T\wedge n} (X_{k+1} - X_k) = X_0 + \sum\limits_{0\leq k} I_{\{k<T\wedge n\}}(X_{k+1} - X_k)$. Comme $I_{\{k<T\wedge n\}} = I_{\{k<T\}} I_{\{k<n\}}$ et $I_{\{k<n\}} = 1$ si $k = 0, 1, \ldots, n-1$, et 0 si $k \geq n$, on en déduit la représentation (8.1.3). □

La variable aléatoire X_T est appelée *variable aléatoire terminale du processus arrêté* $X^{|T}$. Elle est définie chaque fois que le temps d'arrêt T est fini, ou presque sûrement fini. Cette terminologie se justifie par le fait que, pour presque tout ω, la trajectoire du processus $X^{|T}$ est constante à partir de l'instant $T(\omega)$ et que, sur l'intervalle $[T(\omega), +\infty[$, elle prend la valeur $(X_T)(\omega) = X_{T(\omega)}(\omega)$.

Proposition 8.1.2. *Si le temps d'arrêt* T *est presque sûrement fini, à savoir si* $\mathrm{P}\{T < +\infty\} = 1$, *alors*

$$X_{T\wedge n} \xrightarrow{p.s.} X_T \quad (n \to +\infty).$$

Autrement dit, le processus $X^{|T}$ *tend presque sûrement vers la variable aléatoire terminale.*

Démonstration. Pour presque tout ω on a $T(\omega) < +\infty$. Considérons un tel ω ; alors, pour tout $n \geq T(\omega)$, on a $(X_{T\wedge n})(\omega) = X_{T(\omega)\wedge n}(\omega) = X_{T(\omega)}(\omega) = (X_T)(\omega)$. □

Remarque. Lorsque $T(\omega) = +\infty$, alors, pour tout n, on a : $(X_{T\wedge n})(\omega) = X_n(\omega)$, de sorte que le comportement asymptotique de la suite de terme

général $(X_{T\wedge n})(\omega)$ se confond avec celui de la suite de terme général $X_n(\omega)$. Si, par exemple, la suite $(X_n(\omega))$ est convergente, on pose $X_T(\omega) = \lim_{n\to\infty} X_n(\omega)$. On peut alors prolonger la définition de la variable aléatoire X_T, qui n'est définie *a priori* que sur l'ensemble $\{T < +\infty\}$, en lui donnant, sur $\{T = +\infty\}$, la valeur $\lim_{n\to\infty} X_n$. On obtient alors $X_T = \lim_{n\to\infty} X_{T\wedge n}$.

Cette opération peut être effectuée lorsque $X = (X_n)$ est une martingale telle que $\sup_n \mathbb{E}[|X_n|] < +\infty$. Dans un tel cas, la suite (X_n) est presque sûrement convergente. Si T est un temps d'arrêt quelconque, on a alors $X_T = \lim_{n\to\infty} X_{T\wedge n}$.

Proposition 8.1.3. *Soient $X = (X_n)$ $(n \geq 0)$ une martingale (resp. une sous-martingale, resp. une surmartingale) par rapport à $Y = (Y_n)$ $(n \geq 0)$ et T un temps d'arrêt du processus Y. Alors le processus arrêté $X^{|T} = (X_{T\wedge n})$ $(n \geq 0)$ est encore une martingale (resp. une sous-martingale, resp. une surmartingale) par rapport au processus Y.*

Démonstration. La représentation de $X_{T\wedge n}$, qui apparaît dans le Lemme 8.1.1 (formule (8.1.3)), permet d'écrire :

$$X_{T\wedge(n+1)} - X_{T\wedge n} = I_{\{n<T\}}(X_{n+1} - X_n),$$

d'où

$$\begin{aligned}\mathbb{E}[X_{T\wedge(n+1)} - X_{T\wedge n} \,|\, Y_0, \ldots, Y_n] &= \mathbb{E}[I_{\{n<T\}}(X_{n+1} - X_n) \,|\, Y_0, \ldots, Y_n] \\ &= I_{\{n<T\}}\, \mathbb{E}[X_{n+1} - X_n \,|\, Y_0, \ldots, Y_n],\end{aligned}$$

puisque l'indicatrice $I_{\{n<T\}}$ est une fonction du vecteur $(Y_0, \ldots, Y_n)$. Or, si X est une martingale (resp. une sous-martingale, resp. une surmartingale), l'expression $\mathbb{E}[X_{n+1} - X_n \,|\, Y_0, \ldots, Y_n]$ est nulle (resp. positive, resp. négative ou nulle). Comme l'indicatrice $I_{\{n<T\}}$ est évidemment positive, on vérifie bien le fait que $\mathbb{E}[X_{T\wedge(n+1)} - X_{T\wedge n} \,|\, Y_0, \ldots, Y_n]$ est nulle (resp. positive, resp. négative ou nulle). □

8.2 LE THÉORÈME D'ARRÊT POUR UN TEMPS D'ARRÊT BORNÉ

Il s'agit d'évaluer l'espérance mathématique de X_T, lorsque X est une martingale (resp. une sous-martingale, resp. une surmartingale). Nous établissons d'abord le résultat, lorsque T est *borné.*

Théorème 8.2.1. *Soient $X = (X_n)$ $(n \geq 0)$ une martingale (resp. une sous-martingale, resp. une surmartingale) par rapport à un processus $Y = (Y_n)$ $(n \geq 0)$ et T un temps d'arrêt borné adapté à Y. Alors*

$$\begin{gathered}\mathbb{E}[X_T] = \mathbb{E}[X_0] \\ (\textit{resp. } \mathbb{E}[X_T] \geq \mathbb{E}[X_0], \\ \textit{resp. } \mathbb{E}[X_T] \leq \mathbb{E}[X_0]).\end{gathered}$$

Démonstration. Comme T est borné, il existe un entier $M \geq 1$ tel que $0 \leq T \leq M$, d'où $X_{T\wedge M} = X_T$. D'après (8.1.3), on peut donc écrire :

$$X_T = X_{T\wedge M} = X_0 + \sum_{k=0}^{M-1} I_{\{k<T\}}\,(X_{k+1} - X_k).$$

Or, pour chaque $k = 0, 1, \ldots, M-1$, on a :

$$\mathbb{E}[I_{\{k<T\}}\,(X_{k+1} - X_k)] = \mathbb{E}[\mathbb{E}[\,I_{\{k<T\}}\,(X_{k+1} - X_k)\,|\,Y_0, \ldots, Y_k]\,],$$

d'après la Règle 5 sur l'espérance conditionnelle (*cf.* § 1.13). Comme l'indicatrice $I_{\{k<T\}}$ est une fonction du vecteur $(Y_0, \ldots, Y_k)$, on peut encore écrire :

$$\mathbb{E}[I_{\{k<T\}}\,(X_{k+1} - X_k)] = \mathbb{E}[\,I_{\{k<T\}}\,\mathbb{E}[\,X_{k+1} - X_k\,|\,Y_0, \ldots, Y_k]\,].$$

Maintenant, si X est une martingale (resp. une sous-martingale, resp. une surmartingale), on a

$$\mathbb{E}[\,X_{k+1} - X_k\,|\,Y_0, \ldots, Y_k] = 0 \text{ (resp. } \geq 0,\ \text{resp. } \leq 0),$$

d'où

$$\mathbb{E}[\,I_{\{k<T\}}\,(X_{k+1} - X_k)] = 0 \text{ (resp. } \geq 0,\ \text{resp. } \leq 0).$$

Par suite,

$$\begin{aligned}\mathbb{E}[X_T] &= \mathbb{E}[X_0] + \sum_{k=0}^{M-1} \mathbb{E}[\,I_{\{k<T\}}\,(X_{k+1} - X_k)] \\ &= \mathbb{E}[X_0] \text{ (resp. } \geq \mathbb{E}[X_0],\ \text{resp. } \leq \mathbb{E}[X_0]). \quad \square\end{aligned}$$

8.3 LE THÉORÈME D'ARRÊT POUR LES MARTINGALES DOMINÉES

Il s'agit toujours d'arriver à la conclusion $\mathbb{E}[X_T] = \mathbb{E}[X_0]$, lorsque X est une martingale et T un temps d'arrêt, mais que ce dernier n'est plus nécessairement borné. On le suppose, toutefois, presque sûrement fini. On voit qu'avec une hypothèse supplémentaire sur le processus arrêté $(X_{T\wedge n})$ (condition (b) ci-dessous), on peut encore arriver à la conclusion souhaitée.

Théorème 8.3.1. *Soient $X = (X_n)$ $(n \geq 0)$ une martingale par rapport à un processus $Y = (Y_n)$ $(n \geq 0)$ et T un temps d'arrêt adapté à Y. On suppose, de plus, que*

(a) $\mathrm{P}\{T < +\infty\} = 1$;
(b) $\mathbb{E}[\sup_{n\geq 0} |X_{T\wedge n}|\,] < +\infty$.

Alors $\mathbb{E}[\,|X_T|\,] < +\infty$ *et* $\mathbb{E}[X_T] = \mathbb{E}[X_0]$.

Démonstration. Comme $T\wedge n$ est un temps d'arrêt *borné*, le Théorème 8.2.1 appliqué au processus arrêté $(X_{T\wedge n})$ implique $\mathbb{E}[X_{T\wedge n}] = \mathbb{E}[X_0]$. Pour démontrer le théorème, il suffit de prouver que $\mathbb{E}[X_{T\wedge n}] \to \mathbb{E}[X_T]$ et que X_T est intégrable.

Or, d'après la Proposition 8.1.2, la condition (a) implique que l'on a : $X_{T\wedge n} \xrightarrow{p.s.} X_T$ $(n \to +\infty)$. En général, ce résultat n'implique pas nécessairement que l'on a : $\mathbb{E}[X_{T\wedge n}] \to \mathbb{E}[X_T]$. Ceci est pourtant vrai ici, mais pour des raisons bien précises que l'on détaille ci-après. Posons $Z := \sup_{n\geq 0} |X_{T\wedge n}|$. D'après la condition (b), cette variable aléatoire, évidemment positive, est *intégrable* et naturellement, pour tout $n \geq 0$, on a : $|X_{T\wedge n}| \leq Z$. Les conditions d'application du théorème de convergence dominée de Lebesgue (voir [FF1], chap. 10, Théorème 9.3) sont satisfaites. On peut donc conclure que X_T est intégrable et que $\mathbb{E}[X_{T\wedge n}] \to \mathbb{E}[X_T]$, lorsque n tend vers l'infini. □

Remarque. L'hypothèse (b) est, en particulier, vérifiée si le processus arrêté $X^{\upharpoonright T}$ est *borné*.

Corollaire 8.3.2. *Soient $X = (X_n)$ $(n \geq 0)$ une martingale par rapport à un processus $Y = (Y_n)$ $(n \geq 0)$ et T un temps d'arrêt adapté à Y. On suppose, de plus, que*

(a′) $\mathbb{E}[T] < +\infty$;

(b′) *il existe une constante C telle que pour tout $n < T$ on ait* $\mathbb{E}[\,|X_{n+1} - X_n|\,|Y_0, \dots, Y_n\,] \leq C$.

Alors $\mathbb{E}[\,|X_T|\,] < +\infty$ *et* $\mathbb{E}[X_T] = \mathbb{E}[X_0]$.

Démonstration. Il suffit de vérifier que les conditions (a′) et (b′) de ce corollaire entraînent les conditions (a) et (b) du Théorème 8.3.1. Naturellement (a′) $[\mathbb{E}[T] < +\infty]$ entraîne (a) $[\mathrm{P}\{T < +\infty\} = 1]$.

Posons $Z_0 := |X_0|$ et $Z_{k+1} := |X_{k+1} - X_k|$ pour $k \geq 0$. En utilisant l'expression de $X_{T\wedge n}$ donnée en (8.1.3), on obtient

$$|X_{T\wedge n}| \leq Z_0 + \sum_{0\leq k\leq n-1} I_{\{k<T\}} Z_{k+1} \leq Z_0 + \sum_{0\leq k} I_{\{k<T\}} Z_{k+1},$$

une variable aléatoire que l'on pose égale à Wet qui prend, éventuellement, la valeur$+\infty$. Il en résulte l'inégalité $\sup_{n\geq 0} |X_{T\wedge n}| \leq W$. Pour conclure, il suffit donc de démontrer que W est intégrable.

Or, pour tout $k \geq 0$, on a :

$$\begin{aligned}
\mathbb{E}[I_{\{k<T\}} Z_{k+1}] &= \mathbb{E}[\,\mathbb{E}[\,I_{\{k<T\}} Z_{k+1} \,|\, Y_0, \dots, Y_k]\,] \\
&= \mathbb{E}[\,I_{\{k<T\}} \mathbb{E}[\,Z_{k+1} \,|\, Y_0, \dots, Y_k]\,] \\
&\qquad\qquad [\text{puisque } I_{\{k<T\}} \text{ est une fonction de } (Y_0, \dots, Y_k)] \\
&\leq \mathbb{E}[\,I_{\{k<T\}}\, C] \qquad\qquad [\text{d'après (b}')] \\
&= C\,\mathrm{P}\{T > k\}.
\end{aligned}$$

On en tire :

$$\begin{aligned}
\mathbb{E}[W] = \mathbb{E}[Z_0] + \sum_{0\leq k} \mathbb{E}[I_{\{k<T\}} Z_{k+1}] &\leq \mathbb{E}[Z_0] + C \sum_{0\leq k} \mathrm{P}\{T > k\} \\
&= \mathbb{E}[Z_0] + C\,\mathbb{E}[T] < +\infty. \quad \square
\end{aligned}$$

Remarque. La variable aléatoire W, introduite dans la démonstration du corollaire précédent, majore $|X_T|$. En effet, $X_{T\wedge n} \xrightarrow{p.s.} X_T$, puisque T est presque sûrement fini. Ensuite, pour tout $n \geq 0$, on a : $|X_{T\wedge n}| \leq W$. En faisant tendre n vers l'infini, on obtient bien $|X_T| \leq W$.

Le prochain corollaire est en fait un cas particulier du Corollaire 3.2, qui s'applique dans le cas des suites de sommes partielles de variables aléatoires indépendantes, identiquement distribuées et centrées. Voir Proposition 7.2.5.

Corollaire 8.3.3. *Soient $Y_0 = 0$ la variable aléatoire identiquement nulle et (Y_n) $(n \geq 1)$ une suite de variables aléatoires indépendantes, identiquement distribuées, du premier ordre et centrées. Soit, de plus, T un temps d'arrêt adapté à $Y = (Y_n)$ $(n \geq 0)$ tel que $\mathbb{E}[T] < +\infty$. Pour $n \geq 0$, on pose $X_n := Y_0 + Y_1 + \cdots + Y_n$.*

Alors $\mathbb{E}[\,|X_T|\,] < +\infty$ *et* $\mathbb{E}[X_T] = \mathbb{E}[X_0] = 0$.

Démonstration. D'après la Proposition 7.1.5, le processus $X = (X_n)$ $(n \geq 0)$ est une martingale adaptée au processus Y. Il suffit donc de vérifier la condition (b′) du Corollaire 8.3.2. Or $Y_{n+1} = X_{n+1} - X_n$ est, par hypothèse, indépendant du vecteur $(Y_0, \ldots, Y_n)$. Par conséquent, pour tout $n \geq 0$, on a : $\mathbb{E}[\,|X_{n+1} - X_n|\,|\,Y_0, \ldots, Y_n] = \mathbb{E}[\,|Y_{n+1}|\,] = \mathbb{E}[\,|Y_1|\,] = C < +\infty$. On voit que cette inégalité est même vérifiée pour tout n et non seulement pour $n < T$. □

Corollaire 8.3.4 (Les identités de Wald). *Soient $Y_0 = 0$ la variable aléatoire identiquement nulle et (Y_n) $(n \geq 1)$ une suite de variables aléatoires indépendantes, identiquement distribuées et du premier ordre. On pose $\mathbb{E}[Y_1] = \mu$. Soit, de plus, T un temps d'arrêt adapté à $Y = (Y_n)$ $(n \geq 0)$ tel que $\mathbb{E}[T] < +\infty$. Pour $n \geq 0$, on pose $X_n := Y_0 + Y_1 + \cdots + Y_n$.*

Alors $\mathbb{E}[\,|X_T|\,] < +\infty$ *et*

$$\mathbb{E}[X_T] = \mathbb{E}[T]\,\mathbb{E}[Y_1]. \tag{8.3.1}$$

Si, de plus, les Y_n sont du second ordre, à savoir $\mathbb{E}[Y_1^2] < +\infty$, alors

$$\mathbb{E}[\sum_{k=1}^{T}(Y_k - \mu)^2] = \mathbb{E}[T]\ \mathrm{Var}\,Y_1. \tag{8.3.2}$$

Les formules (8.3.1) *et* (8.3.2) *constituent les deux identités de Wald.*

Démonstration. On applique le Corollaire 8.3.3 à la suite (Y'_n), où $Y'_0 := 0$ et $Y'_n := Y_n - \mu$ $(n \geq 1)$, de sorte que la suite (X'_n) $(n \geq 0)$, où $X'_n := Y'_0 + \cdots + Y'_n = X_n - n\mu$ $(n \geq 0)$ est une martingale. On a alors $X'_T = X_T - T\mu$. D'après le Corollaire 8.3.3, il vient : $\mathbb{E}[\,|X'_T|\,] < +\infty$, d'où $\mathbb{E}[\,|X_T|\,] \leq \mathbb{E}[\,|X'_T|\,] + \mathbb{E}[T]\,|\mu| < +\infty$. De plus, $0 = \mathbb{E}[X'_T] = \mathbb{E}[X_T] - \mathbb{E}[T]\,\mu$.

Pour la seconde identité, on pose $Z_0 := 0$, puis

$$Z_n := \sum_{1 \leq k \leq n} (Y_k - \mu)^2 - n\sigma^2 \quad (n \geq 1),$$

où $\sigma^2 = \mathrm{Var}\,Y_1$. Pour chaque $n \geq 1$, la variable aléatoire Z_n est la somme de n variables aléatoires, indépendantes, identiquement distribuées, du premier

ordre et *centrées*. On applique de nouveau le Corollaire 8.3.3. On a :

$$\mathbb{E}[\,|Z_T|\,] < +\infty \quad \text{et} \quad 0 = \mathbb{E}[Z_T] = \mathbb{E}[\sum_{k=1}^{T}(Y_k - \mu)^2] - \mathbb{E}[T]\ \mathrm{Var}\, Y_1. \quad \square$$

8.4 UNE TROISIÈME VARIANTE DU THÉORÈME D'ARRÊT

Dans le prochain énoncé, on s'impose de nouvelles conditions suffisantes pour toujours arriver à la conclusion $\mathbb{E}[X_T] = \mathbb{E}[X_0]$.

Théorème 8.4.1. *Soient $X = (X_n)$ $(n \geq 0)$ une martingale (resp. une sous-martingale, resp. une surmartingale) par rapport à un processus $Y = (Y_n)$ $(n \geq 0)$ et T un temps d'arrêt adapté à Y. On suppose, de plus, que*

(a) $\mathrm{P}\{T < +\infty\} = 1$;

(b) $\mathbb{E}[\,|X_T|\,] < +\infty$.

(c) $\lim_{n\to+\infty} \mathbb{E}[I_{\{T>n\}} X_n] = 0$ *(resp. ≤ 0, resp. ≥ 0).*

Alors $\mathbb{E}[X_T] = \mathbb{E}[X_0]$ *(resp. ≥ 0, resp. ≤ 0).*

Démonstration. Pour tout $n \geq 0$ on a

$$\begin{aligned} X_T - X_{T\wedge n} &= (X_T - X_{T\wedge n})I_{\{T\leq n\}} + (X_T - X_{T\wedge n})I_{\{T>n\}} \\ &= X_T I_{\{T\leq n\}} - X_T I_{\{T\leq n\}} + X_T I_{\{T>n\}} - X_n I_{\{T>n\}} \end{aligned}$$

et donc

$$\mathbb{E}[X_T] = \mathbb{E}[X_{T\wedge n}] - \mathbb{E}[X_n I_{\{T>n\}}] + \mathbb{E}[X_T I_{\{T>n\}}]. \tag{8.4.1}$$

Appliquons le Théorème 8.2.1 pour le temps d'arrêt *borné* $T \wedge n$. On a : $\mathbb{E}[X_{T\wedge n}] = \mathbb{E}[X_0]$ (resp. $\geq \mathbb{E}[X_0]$, resp. $\leq \mathbb{E}[X_0]$). Ensuite, la condition (c) dit que $\mathbb{E}[X_n I_{\{T>n\}}]$ tend vers 0, lorsque n tend vers l'infini. Enfin, d'après la condition (a) de l'énoncé, on a $I_{\{T>n\}} \xrightarrow{p.s.} 0$, d'où $X_T I_{\{T>n\}} \xrightarrow{p.s.} 0$, lorsque n tend vers l'infini. Par ailleurs, pour tout $n \geq 0$, on a $|X_T I_{\{T>n\}}| \leq |X_T|$ avec $\mathbb{E}[\,|X_T|\,] < +\infty$ (condition (b)). On peut donc appliquer le théorème de convergence dominée (*cf.* [FF1], chap. 10, Théorème 9.3) et déduire que $\lim_{n\to+\infty} \mathbb{E}[X_T I_{\{T>n\}}] = \mathbb{E}[0] = 0$. Lorsque n vers l'infini dans (8.4.1), on obtient bien la conclusion du théorème. $\square$

Corollaire 8.4.2. *Soient $X = (X_n)$ $(n \geq 0)$ une martingale par rapport à un processus $Y = (Y_n)$ $(n \geq 0)$ et T un temps d'arrêt adapté à Y. On suppose, de plus, que*

(a′) $\mathrm{P}\{T < +\infty\} = 1$;

(b′) *il existe une constante $C < +\infty$ telle que pour tout $n \geq 0$ on ait* $\mathbb{E}[X_{T\wedge n}^2] \leq C$.

Alors $\mathbb{E}[X_T] = \mathbb{E}[X_0]$.

Démonstration. Montrons que les deux conditions (a′) et (b′) entraînent les conditions (a), (b) et (c) du Théorème 8.4.1. D'abord (a′)=(a). Ensuite, (a′) entraîne que l'on a : $X_{T\wedge n}^2 \xrightarrow{p.s.} X_T^2$, lorsque n tend vers l'infini. Ce dernier

résultat, combiné avec la condition (b′), permet de faire usage du Lemme de Fatou (*cf.* [FF1], chap. 10, Corollaire du Théorème 9.2). On conclut que $\mathbb{E}[X_T^2] \le C$, d'où $\mathbb{E}[\,|X_T|\,] \le \sqrt{\mathbb{E}[X_T^2]} \le \sqrt{C} < +\infty$, qui est la condition (b). Enfin, $(\mathbb{E}[X_n I_{\{T>n\}}])^2 = (\mathbb{E}[X_{T\wedge n} I_{\{T>n\}}])^2 \le \mathbb{E}[X_{T\wedge n}^2]\,\mathbb{E}[I_{\{T>n\}}^2]$, par l'inégalité de Schwarz, soit $(\mathbb{E}[X_n I_{\{T>n\}}])^2 \le C\,\mathrm{P}\{T > n\}$. Comme la condition (a′) entraîne $\lim_n \mathrm{P}\{T > n\} = 0$, la condition (c) est bien vérifiée. □

8.5 LE THÉORÈME D'ARRÊT POUR LES MARTINGALES UNIFORMÉMENT INTÉGRABLES

Il s'agit toujours de trouver des conditions *suffisantes* pour que l'on puisse conclure que la variable X_T, où (X_n) est une martingale et T un temps d'arrêt, a même espérance que la variable X_0. Dans ce dernier paragraphe, c'est la notion d'*intégrabilité uniforme* qui est utilisée de façon cruciale.

Définition 8.5.1. Soit (X_n) $(n \ge 0)$ une suite de variables aléatoires définies sur un même espace probabilisé. On dit que cette suite est *uniformément intégrable* si l'on a :

$$\lim_{c\to 0} \sup_n \mathbb{E}[I_{\{|X_n|>c\}}|X_n|\,] = 0.$$

Comme démontré dans le lemme suivant, cette condition d'intégrabilité uniforme implique que non seulement les X_n sont intégrables, mais que les espérances mathématiques des $|X_n|$ sont bornées par une même constante.

Lemme 8.5.1. *Si la suite (X_n) $(n \ge 0)$ est uniformément intégrable, alors elle est bornée dans L^1, c'est-à-dire il existe une constante $M < +\infty$ telle que, pour tout $n \ge 0$, on ait $\mathbb{E}[\,|X_n|\,] \le M < +\infty$.*

Démonstration. Écrivons $|X_n| = I_{\{|X_n|>c\}}|X_n| + I_{\{|X_n|\le c\}}|X_n| = A_n + B_n$, d'où $\mathbb{E}[\,|X_n|\,] = \mathbb{E}[A_n] + \mathbb{E}[B_n]$. En vertu de l'intégrabilité uniforme, quelque soit $\varepsilon > 0$ il existe une constante $c(\varepsilon)$ telle que pour tout $n \ge 0$ on ait $\mathbb{E}[A_n] \le \varepsilon$. Or, pour cette valeur $c(\varepsilon)$ et pour tout $n \ge 0$, on a l'inégalité $\mathbb{E}[B_n] \le c(\varepsilon)$. On peut donc prendre $M = \varepsilon + c(\varepsilon)$. □

Lemme 8.5.2. *Soient T une variable aléatoire à valeurs dans $\mathbb{N} \cup \{+\infty\}$, supposée presque sûrement finie et (X_n) $(n \ge 0)$ une suite uniformément intégrable de variables aléatoires définies sur le même espace probabilisé que T. Alors*

$$\lim_{n\to+\infty} \mathbb{E}[I_{\{T>n\}}|X_n|\,] = 0. \tag{8.5.1}$$

Démonstration. Pour tout $n \ge 0$ et tout $c > 0$, on a

$$I_{\{T>n\}} = I_{\{T>n\}} I_{\{|X_n|\le c\}} + I_{\{T>n\}} I_{\{|X_n|>c\}}\,;$$

d'où

$$\begin{aligned}\mathbb{E}[I_{\{T>n\}}|X_n|\,] &= \mathbb{E}[I_{\{T>n\}} I_{\{|X_n|\le c\}}|X_n|\,] + \mathbb{E}[I_{\{T>n\}} I_{\{|X_n|>c\}}|X_n|\,]\\ &= A_n + B_n.\end{aligned}$$

Or, pour tout $n \geq 0$, on a $B_n \leq \mathbb{E}[I_{\{|X_n|>c\}}|X_n|\,]$. De plus, d'après l'hypothèse d'intégrabilité uniforme, quelque soit $\varepsilon > 0$, il existe une constante $c(\varepsilon) > 0$ telle que pour tout $n \geq 0$ on ait $B_n \leq \varepsilon$. Prenons alors $c := c(\varepsilon)$ dans la définition de A_n. Il vient : $A_n = \mathbb{E}[I_{\{T>n\}}I_{\{|X_n|\leq c\}}|X_n|\,] \leq c\,\mathrm{P}\{T > n\}$. Comme T est presque sûrement finie, on a $A_n \to 0$ et donc $A_n + B_n \to 0$, lorsque n tend vers l'infini. □

On peut alors énoncer cette dernière variante du théorème d'arrêt.

Théorème 8.5.3. *Soient $X = (X_n)$ $(n \geq 0)$ une martingale adaptée au processus $Y = (Y_n)$ $(n \geq 0)$ et T un temps d'arrêt adapté à Y. On suppose que :*

(1) $\mathrm{P}\{T < +\infty\} = 1$, *i.e. le temps d'arrêt T est presque sûrement fini;*

(2) *la suite (X_n) $(n \geq 0)$ est uniformément intégrable.*

Alors $\mathbb{E}[\,|X_T|\,] < +\infty$ *et* $\mathbb{E}[X_T] = \mathbb{E}[X_0]$.

Démonstration. D'après le Lemme 8.5.1, pour tout $n \geq 0$, on a $\mathbb{E}[\,|X_n|\,] \leq M$, donc aussi $\mathbb{E}[\,|X_{T\wedge n}|\,] \leq M$. D'autre part, puisque T est presque sûrement fini, on a $X_{T\wedge n} \xrightarrow{p.s.} X_T$, d'où $|X_{T\wedge n}| \xrightarrow{p.s.} |X_T|$, lorsque n tend vers l'infini. Le Lemme de Fatou (*cf.*, par exemple, [FF1], chap. 10, Corollaire du Théorème 9.2) permet de conclure que $\mathbb{E}[\,|X_T|\,] \leq M < +\infty$. Les conditions (a), (b) et (c) du Théorème 8.4.1 sont vérifiées, puisque la condition (c) de ce théorème n'est autre que (8.5.1). On a donc bien $\mathbb{E}[X_T] = \mathbb{E}[X_0]$. □

8.6 COMPLÉMENTS ET EXERCICES

8.1 Précisions sur le Théorème 8.2.1

Soient $X = (X_n)$ $(n \geq 0)$ une martingale (resp. une sous-martingale, resp. une surmartingale) adaptée à un processus $Y = (Y_n)$ $(n \geq 0)$ et T un temps d'arrêt *borné* adapté à Y et vérifiant $0 \leq T \leq M$ pour un certain entier $M \geq 1$. Alors $\mathbb{E}[X_0] = \mathbb{E}[X_T] = \mathbb{E}[X_M]$ si X est une martingale, $\mathbb{E}[X_0] \leq \mathbb{E}[X_T] \leq \mathbb{E}[X_M]$ si X est une sous-martingale, $\mathbb{E}[X_0] \geq \mathbb{E}[X_T] \geq \mathbb{E}[X_M]$ si X est une surmartingale.

Pour le processus arrêté $X^{|T} = (X_{T\wedge n})$ $(n \geq 0)$, on a les relations : $\mathbb{E}[X_0] = \mathbb{E}[X_{T\wedge n}] = \mathbb{E}[X_n]$ si X est une martingale, $\mathbb{E}[X_0] \leq \mathbb{E}[X_{T\wedge n}] \leq \mathbb{E}[X_n]$ si X est une sous-martingale et $\mathbb{E}[X_0] \geq \mathbb{E}[X_{T\wedge n}] \geq \mathbb{E}[X_n]$ si X est une surmartingale.

8.2 Calcul de l'espérance mathématique d'une variable géométrique et d'une variable binomiale négative

Soit (Y_n) $(n \geq 1)$ une suite de variables indépendantes, identiquement distribuées, admettant la loi de probabilité : $q\varepsilon_0 + p\varepsilon_1$. Pour $n \geq 1$, on pose $X_n := Y_1 + \cdots + Y_n$. Soit T le temps d'atteinte (*cf.* Exercice 2.1) du processus

(X_n) dans l'état 1. Alors T suit la loi géométrique de paramètre p. Retrouver son espérance mathématique à l'aide de l'identité de Wald.

Pour $r \geq 1$, on pose $T^{(r)} := \inf\{n \geq 1 : X_n = r\}$, de sorte que $T^{(1)} = T$. Alors $T^{(r)}$ suit la loi binomiale négative de paramètres (p, r). Retrouver son espérance mathématique à l'aide de l'identité de Wald.

8.3 Une illusion dissipée

Pierre et Paul jouent à « pile » ou « face » avec la convention suivante : l'enjeu est de 1 euro par partie, Pierre parie toujours sur « pile » et décide de s'arrêter la première fois que son gain total sera de 1 euro. Cette stratégie est-elle bonne ? On considère que les parties sont indépendantes, de même loi, la probabilité d'obtenir « pile » (resp. « face ») étant p (resp. q) avec $p + q = 1$. On désigne par (Y_n) $(n \geq 1)$ la suite des parties, la loi de chaque Y_n étant donnée par $q\varepsilon_{-1} + p\varepsilon_{+1}$. Pour $n \geq 1$, la somme $X_n = Y_1 + \cdots + Y_n$ représente le gain total de Pierre après la $n^{\text{ième}}$ partie. On pose $X_0 = 0$ et $T := \inf\{n \geq 1 : X_n = 1\}$.

(A) On suppose $p = q = 1/2$.

(a) Montrer que l'on a $\mathbb{E}[T] = +\infty$.

(b) Comparer la convergence des martingales (X_n) et $(X_n^{|T})$ sur $\{T = +\infty\}$. En déduire que l'on a $\mathrm{P}\{T < \infty\} = 1$, c'est-à-dire que Pierre gagne à coup sûr. [Pour établir la question (b), on admet le résultat suivant : si $U = (U_n)$ est une martingale et si pour tout $n \geq 1$, on a $U_n \leq V$, où V est une variable aléatoire satisfaisant $\mathbb{E}[\,|V|\,] < +\infty$, alors la suite (U_n) converge presque sûrement.]

(c) On peut aussi établir que $\mathrm{P}\{T = +\infty\} = 0$ directement, en évaluant la probabilité de l'évènement $\{T = n\}$ (voir, par exemple, [FF1], chap. 4, Lemme 5.2). Pour chaque $n \geq 0$, on trouve :

$$\mathrm{P}\{T = 2n+1\} = \frac{1}{2n+1}\binom{2n+1}{n+1}\frac{1}{2^{2n+1}}.$$

Alors $\mathrm{P}\{T < +\infty\} = \sum_{n \geq 0} \mathrm{P}\{T = 2n+1\} = 1$.

(B) On suppose $0 < p < q$. Alors $\mathrm{P}\{T < +\infty\} < 1$, d'où $\mathbb{E}[T] = +\infty$. [Pour $n \geq 1$, on pose $Z_n := (q/p)^{X_n}$ et $Z_0 := 1$; la suite (Z_n) étant une martingale, on applique le théorème d'arrêt pour le temps d'arrêt *borné* $T \wedge n$.]

(C) On suppose $0 < q < p$. Alors $\mathrm{P}\{T < +\infty\} = 1$ et $\mathbb{E}[T] = 1/(p-q)$. [Pour le premier résultat, utiliser la loi forte des grands nombres avec la suite (Y_n). Pour le second résultat, appliquer le Corollaire 3.4 avec le temps d'arrêt $T \wedge n$, pour démontrer : $\mathbb{E}[T \wedge n] \leq 1/(p-q)$. La conclusion $\mathbb{E}[T] = 1/(p-q)$ s'obtient alors par application du Lemme de Fatou (*cf.* [FF1], chap. 10, Corollaire du Théorème 9.2).]

Chapitre 9

Problèmes de ruine

On se propose dans ce chapitre de traiter deux exemples de problèmes de ruine à l'aide de la théorie des martingales développée dans les chapitres précédents. Il s'agit tout d'abord du problème de la ruine d'une compagnie d'assurance, et ensuite du problème classique de la ruine du joueur.

9.1 RUINE D'UNE COMPAGNIE D'ASSURANCE

Les sinistres susceptibles de se produire et devant être couverts par une compagnie d'assurance ont un double aléa : d'abord, l'instant où ils se produisent, ensuite, le montant de dédommagement qu'il faudra débourser, autrement dit, le *coût* du sinistre. On désigne par X une variable aléatoire, à valeurs strictement positives, représentant ce coût. On suppose qu'elle est du second ordre et on pose $\mu = \mathbb{E}[X]$, $\sigma^2 = \operatorname{Var} X < +\infty$.

On se donne, ensuite,

(a) une suite (X_n) $(n \geq 1)$ de variables aléatoires indépendantes, identiquement distribuées comme X ;

(b) un processus de Poisson $\{N(t) : t \geq 0\}$, de densité $\lambda > 0$, indépendant de la suite (X_n) $(n \geq 1)$. Le paramètre λ est le nombre moyen de sinistres par unité de temps.

La somme aléatoire $Y_1 := \sum_{i=1}^{N(1)} X_i$ est alors le *coût des sinistres par unité de temps* et d'après les identités de Wald, on a :

$$\mu_1 := \mathbb{E}[Y_1] = \mathbb{E}[N(1)]\,\mathbb{E}[X] = \lambda\,\mu;$$

puis,

$$\sigma_1^2 := \operatorname{Var} Y_1 = \mathbb{E}[N(1)]\,\mathbb{E}[X^2] = \lambda(\mu^2 + \sigma^2).$$

On considère alors une suite (Y_n) $(n \ge 1)$ de variables aléatoires indépendantes, identiquement distribuées comme Y_1. La somme $C(n) := \sum_{k=1}^{n} Y_k$ est alors le *coût total des sinistres enregistrés dans l'intervalle* $[0, n]$ $(n \ge 1)$. On introduit encore deux autres notions :

(c) *la disponibilité à la date* n : $D(n) := R + P\,n$, où R est la *réserve* et P est le *taux d'entrée des primes* par unité de temps.

(d) *la trésorerie à la date* n :

$$D(n) - C(n) = R + P\,n - C(n) = R + \sum_{k=1}^{n} (P - Y_k).$$

On introduit le temps d'arrêt : $T := \inf\{n \ge 1 : D(n) - C(n) \le 0\}$; c'est le premier instant où la trésorerie de la compagnie est négative ou nulle. La probabilité $\mathrm{P}\{T < +\infty\}$ est la *probabilité de ruine* de la compagnie. Faisons l'hypothèse que la loi de Y_1 est la loi normale $N(\mu_1, \sigma_1)$, où μ_1 est très petit comparé à P. Nous nous proposons de majorer cette probabilité de ruine.

Proposition 9.1.1. *On a l'inégalité :*

$$\mathrm{P}\{T < +\infty\} \le \exp\Bigl(-\frac{2(P - \mu_1)}{\sigma_1^2} R\Bigr) = \exp\Bigl(-\frac{2R(P - \lambda\mu)}{\lambda(\mu^2 + \sigma^2)}\Bigr). \qquad (9.1.1)$$

Nous donnons deux démonstrations de cette Proposition, qui font appel à deux méthodes différentes. Il est remarquable que ces deux méthodes conduisent à la *même* majoration.

9.1.1 Majoration par la méthode des temps d'arrêt

Pour $k \ge 1$, posons $Z_k := P - Y_k$, de sorte que

$$D(n) - C(n) = R + \sum_{k=1}^{n} Z_k \quad (n \ge 1).$$

Comme P est une constante, la suite (Z_n) $(n \ge 1)$ est une suite de variables aléatoires indépendantes, identiquement distribuées comme $Z_1 = P - Y_1$, donc de loi $N(P - \mu_1, \sigma_1)$. Or, la fonction génératrice des moments de Z_1, dont l'expression est donnée par

$$g(u) = \mathbb{E}[e^{uZ_1}] = \exp\Bigl(u(P - \mu_1) + \frac{u^2\sigma_1^2}{2}\Bigr),$$

prend la valeur 1, lorsque $u = u_0 := -2(P - \mu_1)/\sigma_1^2 < 0$. D'après la Proposition 7.3.1, on peut former la *martingale de Wald* associée à la suite de

terme général $D(n) - C(n)$. Cette martingale est donnée par :

$$\begin{cases} W_0(u) = e^{uR}, \\ W_n(u) = e^{uR}\,\dfrac{e^{u(Z_1+\cdots+Z_n)}}{g(u)^n} = \dfrac{e^{u(D(n)-C(n))}}{g(u)^n} \quad (n \geq 1). \end{cases}$$

Pour $u = u_0 = -\dfrac{2(P-\mu_1)}{\sigma_1^2}$, on obtient la martingale :

$$\begin{cases} W_0 := \exp\Big(-\dfrac{2(P-\mu_1)}{\sigma_1^2}R\Big), \\ W_n := \exp\Big(-\dfrac{2(P-\mu_1)}{\sigma_1^2}(D(n)-C(n))\Big) \quad (n \geq 1). \end{cases}$$

L'instant aléatoire $T \wedge m$ ($m \geq 1$) est évidemment un temps d'arrêt *borné* pour la martingale (W_n) ($n \geq 0$). On peut donc appliquer le Théorème d'arrêt 8.2.1

$$\mathbb{E}[W_0] = \mathbb{E}[W_{T\wedge m}];$$

soit

$$\begin{aligned} \exp\Big(-\frac{2(P-\mu_1)}{\sigma_1^2}R\Big) &= \mathbb{E}[W_T I_{\{T\leq m\}}] + \mathbb{E}[W_m I_{\{T>m\}}] \\ &\geq \mathbb{E}[W_T I_{\{T\leq m\}}]. \end{aligned}$$

Or, la trésorerie $D\circ T - C\circ T$ à l'instant T est évidemment négative ou nulle. D'où $W_T \geq 1$ et ainsi

$$\exp\Big(-\frac{2(P-\mu_1)}{\sigma_1^2}R\Big) \geq \mathbb{E}[I_{\{T\leq m\}}] = \mathrm{P}\{T \leq m\}.$$

L'inégalité étant satisfaite pour tout $m \geq 1$, on obtient la majoration (9.1.1), en faisant tendre m vers l'infini. □

9.1.2 Majoration à l'aide des inégalités maximales

Dans cette méthode, on fait appel à l'inégalité suivante, démontrée dans le chapitre 10 (*cf.* Théorème 10.2.1), qui dit que si (X_n) ($n \geq 0$) est une sous-martingale *positive*, alors pour tout $\lambda > 0$, on a la majoration

$$\lambda\,\mathrm{P}\{\sup_{0\leq k\leq n} X_k > \lambda\} \leq \mathbb{E}[X_n]. \tag{9.1.2}$$

Puisque Y_1 suit une loi normale $N(\mu_1, \sigma_1)$, le coût total $C(n) = \sum_{k=1}^{n} Y_k$ dans l'intervalle $[0, n]$ peut s'écrire $C(n) = \mu_1 n + \sigma_1 U_n$, où U_n est une variable aléatoire normale, centrée, d'écart-type $\sqrt{n}$, donc de fonction génératrice des moments $g(u) = e^{u^2 n/2}$. D'après la Proposition 7.3.1, la suite (X_n), où $X_0 := 1$ et

$$X_n := \frac{e^{uU_n}}{e^{u^2 n/2}} = e^{u(U_n - un/2)} \quad (n \geq 1), \tag{9.1.3}$$

est la martingale de Wald associée à (U_n). [Il faudrait dire : associée à la suite $((Y_n-\mu_1)/\sigma_1)$.] Comme les deux évènements

$$\{\sup_{0\le k\le n}(U_k-uk/2)\ge b\} \text{ et } \{\sup_{0\le k\le n}X_k\ge e^{ub}\}$$

sont équivalents et que $\mathbb{E}[X_k]=1$, la majoration (9.1.2) avec $u=:2a$ et $\lambda:=e^{2ab}$ fournit l'inégalité :

$$\mathrm{P}\{\sup_{0\le k\le n}(U_k-ak)\ge b\}\le e^{-2ab}. \tag{9.1.4}$$

Prenons $n\ge 1$; l'évènement "la compagnie est ruinée durant l'intervalle $[0,n]$" est l'évènement $\{\inf_{0\le k\le n}(D(k)-C(k))\le 0\}$. Comme $D(k)-C(k)=R+(P-\mu_1)k-\sigma_1U_k$, il s'exprime encore comme

$$\begin{aligned}\{R-\sup_{0\le k\le n}(\sigma_1U_k-(P-\mu_1)k)\le 0\}&=\{\sup_{0\le k\le n}(\sigma_1U_k-(P-\mu_1)k)\ge R\}\\&=\{\sup_{0\le k\le n}(U_k-ak)\ge b\},\end{aligned}$$

en posant $a:=(P-\mu_1)/\sigma_1$ et $b:=R/\sigma_1$. Conservant ces deux valeurs de a et b, la majoration (9.1.4) fournit alors l'inégalité :

$$\mathrm{P}\{\inf_{0\le k\le n}(D(k)-C(k))\le 0\}\le e^{-2ab}.$$

Comme cette inégalité est uniforme en n, on en déduit la majoration suivante de la probabilité de ruine de la compagnie

$$\mathrm{P}\{\inf_{n\ge 0}(D(n)-C(n))\le 0\}\le e^{-2ab}=\exp\Big(-\frac{2(P-\mu_1)}{\sigma_1^2}R\Big),$$

une majoration qui est identique à celle trouvée par la première méthode.

9.2 LE PROBLÈME DE LA RUINE DES JOUEURS

La théorie des chaînes de Markov permet de traiter ce problème, une fois résolues certaines équations de récurrence. Il existe une autre méthode de résolution faisant appel aux résultats développés jusqu'ici sur les martingales, utilisant notamment le théorème d'arrêt. Nous nous proposons d'exposer cette méthode.

Rappelons que dans le problème de la ruine des joueurs, on suppose que deux joueurs A et B jouent un nombre illimité de parties *indépendantes*, l'enjeu étant de 1 euro par partie : si A gagne la $n^{\text{ième}}$ partie, il donne un euro à B et réciproquement. La probabilité que le joueur A (resp. B) gagne une partie donnée est p (resp. q) et on suppose : $0\le p\le 1$, $p+q=1$. Pour $n\ge 1$, on désigne par Y_n le gain de A à la $n^{\text{ième}}$ partie et l'on pose $Y_0:=0$ et $X_n:=Y_0+Y_1+\cdots+Y_n$ $(n\ge 0)$. On suppose encore que la fortune *initiale* de A est de a euros et celle de B de b euros, où $a,b\ge 1$.

Dans ces conditions, la suite (Y_n) $(n \geq 1)$ est une suite de variables aléatoires, indépendantes, chacune de loi $q\varepsilon_{-1} + p\varepsilon_{+1}$. De plus, X_n représente le *gain* réalisé par A au cours des n premières parties. Après la $n^{\text{ième}}$ partie, la *fortune* de A est $a + X_n$, celle de B de $b - X_n$, tant que ces quantités restent positives. On définit donc un *temps d'arrêt* T du jeu comme étant

$$T := \min\{n \geq 1 : X_n = -a \text{ ou } X_n = +b\}. \tag{9.2.1}$$

La variable aléatoire T est évidemment un temps d'arrêt adapté à (Y_n) $(n \geq 0)$, puisque c'est le *temps d'atteinte* du processus (X_n) $(n \geq 0)$ dans l'ensemble $\{-a, b\}$, qui s'exprime comme une fonction des seules variables Y_0, $Y_1, \ldots, Y_n$.

Proposition 9.2.1. *Quelle que soit la valeur de p $(0 \leq p \leq 1)$, le temps d'arrêt T est presque sûrement fini.*

Démonstration.

(1) Cas $p = q = \frac{1}{2}$. Les variables aléatoires Y_n $(n \geq 1)$ sont alors *centrées.* On peut utiliser la Proposition 7.1.5 pour dire que la suite (X_n) $(n \geq 0)$ est une *martingale* par rapport à (Y_n) $(n \geq 0)$. Or la série de terme général Y_n est non convergente, puisque son terme général ne tend pas vers 0. Considérons alors le processus arrêté $X^{\mid T} = (X_{T\wedge n})$. C'est une martingale (*cf.* Proposition 7.1.3). De plus, pour tout $n \geq 1$, on a : $-a \leq X_{T\wedge n} \leq b$. D'après le Théorème 11.1.1, sur la convergence des martingales, la martingale $X^{\mid T}$ converge presque sûrement. Or, sur l'ensemble $\{T = +\infty\}$, les martingales X et $X^{\mid T}$ coïncident, la seconde converge presque sûrement, mais non la première. Pour lever cette contradiction, on doit conclure que $\mathrm{P}\{T = +\infty\} = 0$.

(2) Cas $p \neq q$. D'après la loi forte des grands nombres, on a :

$$\frac{X_n}{n} = \frac{Y_1 + \cdots + Y_n}{n} \xrightarrow{p.s.} \mathbb{E}[Y_1] = p - q \neq 0.$$

Donc $X_n \xrightarrow{p.s.} +\infty$ si $p - q > 0$ et $X_n \xrightarrow{p.s.} -\infty$ si $p - q < 0$, ce qui suffit pour montrer que $\mathrm{P}\{T < \infty\} = 1$. □

9.3 LA PROBABILITÉ DE RUINE

Le temps d'arrêt T étant presque sûrement fini, on peut considérer la variable aléatoire X_T. C'est le gain (algébrique) de A lorsque le jeu s'arrête. Il ne prend donc, presque sûrement, que les valeurs $-a$ (ruine de A) et b (ruine de B). On pose

$$r_a := \mathrm{P}\{X_T = -a\} \quad \text{et} \quad r_b := \mathrm{P}\{X_T = b\}. \tag{9.3.1}$$

Comme $r_a + r_b = 1$, la loi de X_T est ainsi $r_a\varepsilon_a + (1 - r_a)\varepsilon_b$. L'espérance mathématique de X_T vaut :

$$\mathbb{E}[X_T] = r_a(-a) + (1 - r_a)b. \tag{9.3.2}$$

9.3.1 Cas $p = q = \frac{1}{2}$

La suite (X_n) $(n \geq 0)$ est une martingale. De plus, elle satisfait les conditions d'application du Théorème d'arrêt des martingales dominées (Théorème 8.3.1). D'abord, T est presque sûrement fini, puis $\sup_{n\geq 0} |X_{T\wedge n}| \leq \max\{a, b\}$, d'où $\mathbb{E}[\sup_{n\geq 0} |X_{T\wedge n}|] < +\infty$. Il en résulte $\mathbb{E}[X_T] = \mathbb{E}[X_0] = 0$, d'où, d'après (9.3.2),

$$r_a = \frac{b}{a+b}. \tag{9.3.3}$$

9.3.2 Cas $p \neq q$

Comme $\mathbb{E}[Y_1] = p - q \neq 0$, la suite (X_n) n'est plus une martingale par rapport à (Y_n). On peut cependant construire une autre martingale, qui permettra d'évaluer r_a. Posons, en effet,

$$Z_0 := 1, \quad Z_n := \Big(\frac{q}{p}\Big)^{X_n} \ (n \geq 1). \tag{9.3.4}$$

Lemme 9.3.1. *La suite (Z_n) $(n \geq 0)$ est une martingale par rapport à (Y_n) $(n \geq 0)$. C'est, en fait, une martingale de Wald par rapport à (Y_n) $(n \geq 0)$.*

Démonstration. En effet, on a $g(u) := \mathbb{E}[e^{uY_1}] = pe^u + qe^{-u}$. D'après la Proposition 7.3.1, la suite $Z_0 := 1$, $Z_n := g(u)^{-n}e^{u(Y_1+\cdots+Y_n)}$ $(n \geq 1)$ est une martingale (de Wald) par rapport à (Y_n) $(n \geq 0)$, pourvu que $u \neq 0$. Prenons une valeur de $u \neq 0$ telle que $g(u) = pe^u + qe^{-u} = 1$. En posant $v := e^u$, il s'agit de résoudre l'équation $pv^2 - v + q = 0$. En écartant la solution $v = 1$, on trouve la solution $v = e^u = q/p$. Pour $n \geq 1$, on peut donc prendre : $Z_n = (q/p)^{X_n}$. □

Comme $Z_T = (q/p)^{X_T}$, la loi de Z_T est donnée par

$$r_a\, \varepsilon_{(q/p)^{-a}} + (1 - r_a)\, \varepsilon_{(q/p)^b}$$

et son espérance mathématique par

$$\mathbb{E}[Z_T] = r_a\Big(\frac{q}{p}\Big)^{-a} + (1 - r_a)\Big(\frac{q}{p}\Big)^{b}. \tag{9.3.5}$$

Une nouvelle fois, le théorème d'arrêt fournit la valeur de la probabilité de ruine r_a. Vérifions, en effet, que les hypothèses du Théorème d'arrêt 8.3.1 sont satisfaites pour la martingale (Z_n) $(n \geq 0)$. D'abord, $\mathrm{P}\{T < +\infty\} = 1$, d'après la Proposition 9.2.1. Ensuite, $Z_{T\wedge n} = (q/p)^{X_{T\wedge n}}$ est borné pour $n \geq 1$, puisque $X_{T\wedge n}$ ne prend qu'un nombre fini de valeurs. Il en résulte que $\mathbb{E}[|\sup_{n\geq 0} |Z_{T\wedge n}|] < +\infty$. D'après le théorème juste cité, on en conclut que $\mathbb{E}[Z_T] = \mathbb{E}[Z_0] = 1$, d'où, d'après (9.3.5), $r_a(q/p)^{-a} + (1 - r_a)(q/p)^b = 1$ et donc

$$r_a = \frac{1 - \Big(\frac{q}{p}\Big)^b}{\Big(\frac{q}{p}\Big)^{-a} - \Big(\frac{q}{p}\Big)^b} = \frac{1 - \Big(\frac{p}{q}\Big)^b}{1 - \Big(\frac{p}{q}\Big)^{a+b}}. \tag{9.3.6}$$

9.3.3 Le jeu avec un adversaire infiniment riche

Supposons que la fortune du joueur B est infiniment grande, mais que celle du joueur A est toujours $a < +\infty$. Dans le cas $p = q = \frac{1}{2}$, la formule (9.3.3) entraîne que lorsque b tend vers l'infini, r_a tend vers 1. La ruine de A est *certaine*.

Supposons $p < q$. Lorsque b tend vers l'infini, la formule (9.3.6) entraîne que r_a tend vers 1. Même conclusion.

Supposons $p > q$. Lorsque b tend vers l'infini, la formule (9.3.6) entraîne que r_a tend vers $(q/p)^a$. Dans ce cas, la probabilité de ruine de A est d'autant plus faible que sa fortune initiale a est plus grande.

9.4 LA DURÉE DU JEU

Conservons les notations des paragraphes 2 et 3. Nous nous proposons d'évaluer la durée moyenne du jeu, $\mathbb{E}[T]$, jusqu'à la ruine de l'un des joueurs.

9.4.1 Cas $p = q = \frac{1}{2}$

Supposons que l'on sache que $\mathbb{E}[T]$ est finie. La formule de Wald donne $\mathbb{E}[X_T] = \mathbb{E}[T]\,\mathbb{E}[Y_1]$, mais comme $\mathbb{E}[X_T] = \mathbb{E}[Y_1] = 0$, on ne peut rien conclure sur $\mathbb{E}[T]$. Or, nous savons (*cf.* Exercice 7.1) que la suite (Z_n) $(n \geq 0)$, où $Z_0 := 0$ et $Z_n := X_n^2 - n$ est une martingale par rapport à (Y_n) $(n \geq 0)$. Appliquons-lui le théorème d'arrêt pour le temps d'arrêt *borné* $T \wedge n$ (Théorème 8.2.1). Il vient :

$$0 = \mathbb{E}[Z_0] = \mathbb{E}[Z_{T\wedge n}] = \mathbb{E}[X_{T\wedge n}^2] - \mathbb{E}[T \wedge n]. \tag{9.4.1}$$

Proposition 9.4.1. *On a :* $\mathbb{E}[X_T^2] = \mathbb{E}[T]$.

Démonstration. Comme T est presque sûrement fini, $X_{T\wedge n} \xrightarrow{p.s.} X_T$, d'après la Proposition 8.1.2 ; donc aussi $X_{T\wedge n}^2 \xrightarrow{p.s.} X_T^2$. Comme la suite $(X_{T\wedge n})$ $(n \geq 0)$ est bornée, la suite $(X_{T\wedge n}^2)$ $(n \geq 0)$ l'est aussi. D'après le Théorème de convergence dominée de Lebesgue, $\mathbb{E}[X_{T\wedge n}^2] \to \mathbb{E}[X_T^2]$. D'autre part, comme T est presque sûrement fini, la suite $(T \wedge n)$ est une suite monotone croissante tendant presque sûrement vers une limite finie T. D'après le Théorème de convergence monotone, il vient $\mathbb{E}[T\wedge n] \to \mathbb{E}[T]$. Il suffit donc de faire tendre n vers l'infini dans la formule (9.4.1) pour conclure. □

D'après cette dernière proposition, on a donc : $\mathbb{E}[T] = \mathbb{E}[X_T^2] = r_a(-a)^2 + (1 - r_a)b^2 = \dfrac{b}{a+b}a^2 + \dfrac{a}{a+b}b^2 = ab$. *La durée moyenne du jeu est égale au produit des fortunes initiales des deux joueurs :*

$$\mathbb{E}[T] = ab. \tag{9.4.2}$$

Cette durée est considérablement plus longue que ce que l'on pourrait penser. Pour $a = b = 500$ euros, elle est de 250.000. Pour $a = 1$ et $b = 1000$, elle est de mille parties.

9.4.2 Cas $p \neq q$

La formule de Wald donne $\mathbb{E}[X_T] = \mathbb{E}[T]\,\mathbb{E}[Y_1] = \mathbb{E}[T]\,(p-q)$. Comme $p-q \neq 0$, on pourrait en déduire la valeur de $\mathbb{E}[T]$ si l'on savait, *a priori*, qu'elle était finie. On peut s'affranchir de cette hypothèse en procédant comme suit.

La suite (X_n) $(n \geq 0)$ n'est pas une martingale, mais la suite (Z_n) $(n \geq 0)$, où $Z_n := X_n - n\,\mu$, avec $\mu := \mathbb{E}[Y_1] = p-q$, en est une. On lui applique le théorème d'arrêt pour le temps d'arrêt *borné* $T \wedge n$ (*op. cit.*) et l'on obtient :

$$0 = \mathbb{E}[Z_0] = \mathbb{E}[Z_{T\wedge n}] = \mathbb{E}[X_{T\wedge n}] - \mathbb{E}[T \wedge n]\,\mu. \tag{9.4.3}$$

On reprend alors la même démonstration que celle de la Proposition 4.1 pour se convaincre qu'en faisant tendre n vers l'infini dans la formule (9.4.3), on a : $\mathbb{E}[T] = \lim_n \mathbb{E}[T \wedge n] = \lim_n \mathbb{E}[X_{T\wedge n}] = \mathbb{E}[X_T]$. Ceci montre, d'abord, que $\mathbb{E}[T]$ est finie. Ensuite, la relation

$$\mathbb{E}[T] = \frac{1}{\mu}\,\mathbb{E}[X_T] \tag{9.4.4}$$

donne la formule $\mathbb{E}[T] = \dfrac{1}{p-q}(r_a(-a) + (1-r_a)b) = \dfrac{1}{p-q}(b - (a+b)r_a)$, soit, d'après (9.3.6),

$$\mathbb{E}[T] = \frac{b}{p-q} - \frac{a+b}{p-q}\,\frac{1-\left(\dfrac{p}{q}\right)^b}{1-\left(\dfrac{p}{q}\right)^{a+b}}. \tag{9.4.5}$$

Cette formule est invariante dans la transformation $(a,p) \leftrightarrow (b,q)$:

$$\mathbb{E}[T] = \frac{a}{q-p} - \frac{a+b}{q-p}\,\frac{1-\left(\dfrac{q}{p}\right)^a}{1-\left(\dfrac{q}{p}\right)^{a+b}}. \tag{9.4.6}$$

9.4.3 Le jeu avec un adversaire infiniment riche

Il s'agit d'étudier le comportement de la durée du jeu $\mathbb{E}[T]$, exprimée en fonction de a, b, p, q, lorsque a est fixé et b tend vers l'infini.

Dans le cas $p = q = \frac{1}{2}$, la relation (9.4.2) donne : $\mathbb{E}[T] \to +\infty$. Le joueur A joue un jeu équitable avec un adversaire infiniment riche. Sa ruine est certaine, mais il a la consolation d'obliger son adversaire à jouer pendant une durée d'espérance mathématique infinie, avant qu'il soit ruiné !

Lorsque $p < q$, la relation (9.4.5) donne : $\mathbb{E}[T] \to a/(q-p)$. Le joueur A est, à chaque partie, en désavantage contre un adversaire, qui est, en plus, infiniment riche. Sa ruine est certaine et se produit après une durée d'espérance mathématique finie.

Enfin, lorsque $p > q$, la relation (9.4.6) donne : $\mathbb{E}[T] \to +\infty$. Le joueur A a l'avantage à chaque partie ; la probabilité de sa ruine n'est pas certaine (elle est égale à $(q/p)^a$ (*cf.* § 9.3)). La durée du jeu est d'espérance mathématique infinie. Il faut que le joueur ferraille bien et longtemps !

9.5 COMPLÉMENTS ET EXERCICES

9.1 On reprend pour les suites (Y_n) et (X_n) $(n \geq 0)$ les définitions données dans le paragraphe 9.2 pour le problème de la ruine des joueurs. On suppose, de plus $p \neq q$. On démontre directement, sans faire appel au Lemme 3.1, que la suite $Z_0 := 1$, $Z_n := (q/p)^{X_n}$ $(n \geq 1)$ est une martingale par rapport à (Y_n).

9.2 On reprend les mêmes notations que dans les paragraphes 9.2–9.4. On suppose $b \geq 1$ et l'on désigne par $s := a + b$ la fortune totale initiale des deux joueurs. Soit $g(a, s)$ la probabilité pour le joueur A de gagner la somme s. On note $s_{\max}$ la valeur de s qui maximise la probabilité $g(a, s)$. Dans les deux cas $p = q = \frac{1}{2}$ et $p \neq q$, on trouve : $s_{\max} = a + 1$, mais le calcul n'est pas le même!

9.3 « Audentes fortuna juvat »
Toujours dans le problème de la ruine des joueurs, où l'on pose $s := a + b$, il s'agit de savoir si en misant chaque fois $(1/x)$ euros au lieu de 1 euro $(0 < x < 1)$, le joueur A a plus de chances de convertir sa fortune initiale a en s. On convient aussi que le joueur B mise aussi $(1/x)$ à chaque partie. Soit $G(x)$ la probabilité pour le joueur A de gagner la somme s. Dans le cas $p = q = \frac{1}{2}$, on trouve $G(x) = a/s$ et dans le cas $p \neq q$, on trouve $G(x) = (1 - \rho^{ax})/(1 - \rho^{sx})$, en posant $\rho := q/p$. Si $a \geq 1$, $b \geq 1$ et $\rho > 1$, la fonction $G(x)$ est une fonction croissante de la mise $(1/x)$. Le joueur A doit donc choisir la plus grande mise possible, compatible avec sa fortune.

9.4 On considère une suite (Y_n) $(n \geq 0)$ de variables aléatoires, indépendantes, de même loi, celle-ci étant donnée par $q\varepsilon_{-1} + p\varepsilon_{+1}$. Les entiers relatifs $x, c \in \mathbb{Z}$ étant fixés, on pose :

$$S_0^x := x, \quad S_n^x := x + Y_1 + \cdots + Y_n \ (n \geq 1), \quad T_c^x := \inf\{n \geq 0 : S_n^x = c\}.$$

Dans le problème de la ruine des joueurs, on peut interpréter x comme la fortune initiale du joueur A et S_n^x comme sa fortune après la $n^{\text{ième}}$ partie et T_n^x comme le premier instant où sa fortune est égale à c.

(a) Étant donné $d \in \mathbb{Z}$ tel que $c < x < d$, on évalue $\mathrm{P}\{T_c^x < T_d^x\}$, qui est la probabilité, partant de x, d'atteindre c *avant* d. On trouve :

$$\mathrm{P}\{T_c^x < T_d^x\} = \begin{cases} \dfrac{d - x}{d - c}, & \text{si } p = \frac{1}{2}\,; \\ \dfrac{1 - (p/q)^{d-x}}{1 - (p/q)^{d-c}}, & \text{si } p \neq \frac{1}{2}. \end{cases}$$

On vérifie que $\mathrm{P}\{T_c^x < T_d^x\} + \mathrm{P}\{T_d^x < T_c^x\} = 1$.

(b) On pose $\mu_x := \inf\{n \geq 1 : S_n^x = x\}$; c'est le premier instant où la fortune de A est *de nouveau* égale à x. Alors $P\{\mu_x < +\infty\} = 2\min\{p, q\}$.

(c) On pose $M := \sup_{n\geq 0} S_n^0$ et $m := \inf_{n\geq 0} S_n^0$. Ce sont les valeurs extrêmes de la fortune de A au cours du jeu, lorsque sa fortune initiale est 0. Lorsque $p < q$ (resp. $p > q$), la variable aléatoire M (resp. m) suit une loi géométrique. Trouver les expressions. Prolonger les calculs aux cas de $M^x := \sup_{n\geq 0} S_n^x$ et $m^x = \inf_{n\geq 0} S_n^x$.

9.5 Doubler la mise

Soit (X_n) $(n \geq 1)$ une suite de variables aléatoires, indépendantes, toutes de loi : $\frac{1}{2}(\varepsilon_{-1} + \varepsilon_{+1})$. Cette suite représente les parties successives d'un jeu de «pile» ou «face» effectuées par un joueur. Au départ, ce joueur parie 1 euro, s'il gagne, il a retrouvé sa mise et s'arrête. S'il perd, il joue une seconde fois, mais parie cette fois 2 euros ; si le résultat lui est favorable, il empoche sa mise et s'arrête de jouer. Il retrouve donc sa fortune initiale : $-1 + 2 = 1$ euro. S'il perd encore la seconde fois, il joue une troisième fois, mais en pariant $2^2 = 4$ euros. S'il gagne cette troisième partie, il empoche la mise et retrouve ainsi son bien : $-1 - 2 + 4 = 1$, et ainsi de suite... S'il a perdu $(n-1)$ parties, mais gagne la $n^{\text{ième}}$, avec cette stratégie, il a donc perdu : $-1 - 2 - 2^2 - \cdots - 2^{n-1}$, mais comme il parié 2^n la $n^{\text{ième}}$ fois, il retrouve sa fortune : $-1 - 2 - 2^2 - \cdots - 2^{n-1} + 2^n = 1$. Pour $n \geq 1$, on désigne par X_n le gain total (algébrique !) du joueur après la $n^{\text{ième}}$ partie, en convenant que s'il gagne la $k^{\text{ième}}$ partie utilisant cette stratégie, on a : $X_n = 1$ pour tout $k \geq n$. On pose $X_0 := 0$.

(a) Déterminer la loi de X_n et son espérance mathématique.

(b) La suite (X_n) $(n \geq 0)$ est une martingale par rapport à (X_n) $(n \geq 0)$.

(c) On pose : $T := \inf\{n \geq 1 : X_n = 1\}$. C'est donc le nombre de parties que doit jouer le joueur pour retrouver sa fortune en utilisant la stratégie décrite. Alors T est un temps d'arrêt adapté à (X_n) $(n \geq 0)$. Déterminer sa loi de probabilité ; en déduire que $P\{T < +\infty\} = 1$. Montrer que le théorème d'arrêt n'est *pas* vérifié. Quelle est l'hypothèse du Théorème 8.3.1, à savoir le théorème d'arrêt pour les martingales dominées, qui n'est pas satisfaite ?

(d) Si la probabilité d'obtenir «pile» n'est plus $\frac{1}{2}$, mais p ($p \neq \frac{1}{2}$ et que le joueur joue toujours "pile", la suite (X_n) est une sous-martingale ou une surmartingale suivant que $p > \frac{1}{2}$ ou $p < \frac{1}{2}$.

Chapitre 10

Les inégalités maximales

Les inégalités dites *maximales* sont valables pour les martingales et sont fort utiles dans les applications. Elles font l'objet du présent chapitre.

10.1 LES INÉGALITÉS DE BIENAYMÉ-TCHEBYCHEV ET DE KOLMOGOROV

Soit Y_0 la variable aléatoire identiquement nulle et (Y_n) $(n \geq 1)$ une suite de variables aléatoires, indépendantes, identiquement distribuées, centrées et du second ordre. On note $\operatorname{Var} Y_1 = \sigma^2$ la variance de Y_1. On pose, de plus, $S_n := Y_0 + Y_1 + \cdots + Y_n$ $(n \geq 0)$, de sorte que pour tout $n \geq 1$, on a : $\mathbb{E}[S_n] = 0$ et $\operatorname{Var} S_n = n\sigma^2$.

L'*inégalité de Bienaymé-Tchebychev* (*cf.* [FF1], chap. 8, Proposition 8.1) s'écrit :

$$\forall \varepsilon > 0 \quad \varepsilon^2 \, \mathrm{P}\{\, |S_n| > \varepsilon\} \leq \mathbb{E}[S_n^2] = n\sigma^2. \tag{10.1.1}$$

L'*inégalité de Kolmogorov* s'exprime comme suit :

$$\forall \varepsilon > 0 \quad \varepsilon^2 \, \mathrm{P}\{ \max_{0 \leq k \leq n} |S_k| > \varepsilon\} \leq \mathbb{E}[S_n^2] = n\sigma^2. \tag{10.1.2}$$

Il y a des démonstrations élémentaires de l'inégalité (10.1.2). Nous verrons qu'elle est une conséquence de l'une des inégalités maximales que nous allons établir. D'après la Proposition 7.1.5, la suite (S_n) $(n \geq 0)$ est une martingale par rapport à (Y_n) $(n \geq 0)$. D'après l'Exercice 7.10, la suite (S_n^2) $(n \geq 0)$ est alors une *sous-martingale* positive par rapport à (Y_n) $(n \geq 0)$. En fait, l'inégalité (10.1.2) se généralise à *toute* sous-martingale positive.

10.2 LES INÉGALITÉS POUR LES SOUS-MARTINGALES POSITIVES

Théorème 10.2.1. *Soit* (X_n) $(n \geq 0)$ *une sous-martingale positive. Alors*

$$\forall \lambda > 0 \quad \lambda \, \mathrm{P}\{\max_{0 \leq k \leq n} X_k > \lambda\} \leq \mathbb{E}[X_n]. \tag{10.2.1}$$

Démonstration. Posons $M_n := \{\max_{0\leq k\leq n} X_k > \lambda\} = \bigcup_{0\leq k\leq n} \{X_k > \lambda\}$ et introduisons le temps d'arrêt T, qui vaut $\min\{k : 0 \leq k \leq n, X_k > \lambda\}$ sur M_n et n sur M_n^c. C'est bien un temps d'arrêt du processus (X_n) ; de plus, il est *borné* (par n), de sorte qu'on a : $X_{T\wedge n} = X_T$.

Appliquons le théorème d'arrêt pour les temps d'arrêt bornés (Théorème 8.2.1) à la sous-martingale (X_n) $(n \geq 0)$ et au temps d'arrêt $T \wedge n = T$. On a :

$$\begin{aligned}
\mathbb{E}[X_n] &\geq \mathbb{E}[X_{T\wedge n}] = \mathbb{E}[X_T] \\
&= \mathbb{E}[X_T I_{M_n}] + \mathbb{E}[X_T I_{M_n^c}] = \mathbb{E}[X_T I_{M_n}] + \mathbb{E}[X_n I_{M_n^c}] \\
&\geq \mathbb{E}[X_T I_{M_n}] \qquad \text{[puisque la sous-martingale } (X_n) \text{ est positive]} \\
&\geq \lambda \, \mathbb{E}[I_{M_n}] = \lambda \, \mathrm{P}(M_n). \quad \square
\end{aligned}$$

Corollaire 10.2.2. *Soit* (X_n) $(n \geq 0)$ *une martingale. Alors*

$$\forall \lambda > 0 \quad \lambda \, \mathrm{P}\{\max_{0 \leq k \leq n} |X_k| > \lambda\} \leq \mathbb{E}[\,|X_n|\,]. \tag{10.2.2}$$

Démonstration. En effet, si (X_n) est une martingale, alors $(\,|X_n|\,)$ est une sous-martingale positive. $\square$

Corollaire 10.2.3. *Soit* (X_n) $(n \geq 0)$ *une martingale telle que pour tout* $n \geq 0$ *on ait* $\mathbb{E}[X_n^2] < +\infty$. *Alors*

$$\forall \lambda > 0 \quad \lambda^2 \, \mathrm{P}\{\max_{0 \leq k \leq n} |X_k| > \lambda\} \leq \mathbb{E}[\,X_n^2\,]. \tag{10.2.3}$$

Démonstration. Avec les présentes hypothèses, on sait que (X_n^2) $(n \geq 0)$ est une sous-martingale positive. Donc

$$\forall \lambda > 0 \quad \lambda \, \mathrm{P}\{\max_{0 \leq k \leq n} X_k^2 > \lambda\} \leq \mathbb{E}[X_n^2].$$

Avec la substitution $\lambda \leftarrow \lambda^2$, on obtient bien la relation (10.2.3). $\square$

Remarque. Reprenons la martingale (S_n) $(n \geq 0)$ introduite dans le paragraphe 1. Si on lui applique la relation (10.2.3), on obtient l'inégalité de Kolmogorov (10.1.2).

10.3 LES INÉGALITÉS MAXIMALES POUR LES SOUS-MARTINGALES DE SIGNE QUELCONQUE

Théorème 10.3.1. *Soit (X_n) $(n \geq 0)$ une sous-martingale de signe quelconque. Pour $n \geq 0$, soit M_n l'évènement $\{\max\limits_{0\leq k\leq n} X_k > \lambda\}$. Alors*

$$\forall\,\lambda > 0 \quad \lambda\,\mathrm{P}(M_n) \leq \mathbb{E}[\,X_n I_{M_n}\,] \leq \mathbb{E}[X_n^+] \leq \mathbb{E}[\,|X_n|\,]. \qquad (10.3.1)$$

Démonstration. Il suffit de démontrer la première inégalité, les deux autres étant triviales. Reprenons la démonstration du Théorème 10.2.1. Le théorème d'arrêt permet d'écrire $\mathbb{E}[X_n] = \mathbb{E}[X_n I_{M_n}] + \mathbb{E}[X_n I_{M_n^c}] \geq \mathbb{E}[X_T I_{M_n}] + \mathbb{E}[X_n I_{M_n^c}]$, d'où $\mathbb{E}[X_n I_{M_n}] \geq \mathbb{E}[X_T I_{M_n}] > \lambda\,\mathrm{P}(M_n)$. □

Corollaire 10.3.2. *Soit (X_n) $(n \geq 0)$ une martingale. Alors*

$$\forall\,\lambda > 0 \quad \lambda\,\mathrm{P}\{\max_{0\leq k\leq n} |X_k| > \lambda\} \leq \mathbb{E}[\,|X_n|\,I_{\{\max\limits_{0\leq k\leq n} |X_k| > \lambda\}}]. \qquad (10.3.2)$$

Démonstration. En effet, si (X_n) $(n \geq 0)$ est une martingale, alors $(\,|X_n|\,)$ $(n \geq 0)$ est une sous-martingale. □

Corollaire 10.3.3. *Soit (X_n) $(n \geq 0)$ une martingale telle que pour tout $n \geq 0$ on ait $\mathbb{E}[X_n^2] < +\infty$. Alors*

$$\forall\,\lambda > 0 \quad \lambda^2\,\mathrm{P}\{\max_{0\leq k\leq n} |X_k| > \lambda\} \leq \mathbb{E}[\,X_n^2\,I_{\{\max\limits_{0\leq k\leq n} |X_k| > \lambda\}}]. \qquad (10.3.3)$$

Démonstration. En effet, si (X_n) $(n \geq 0)$ est une martingale, alors (X_n^2) $(n \geq 0)$ est une sous-martingale et l'on a :

$$\forall\,\lambda > 0 \quad \lambda\,\mathrm{P}\{\max_{0\leq k\leq n} X_k^2 > \lambda\} \leq \mathbb{E}[\,X_n^2\,I_{\{\max\limits_{0\leq k\leq n} X_k^2 > \lambda\}}].$$

Il suffit de faire la substitution $\lambda \leftarrow \lambda^2$ pour obtenir la relation (10.3.3). □

On sait (*cf.* [FF1], chap. 11, Théorème 4.5) que si X est une variable aléatoire *positive*, son espérance mathématique (dans $[0,+\infty]$) admet aussi l'expression : $\int_0^{+\infty} \mathrm{P}\{X > x\}\,dx$. On a un résultat analogue pour les moments d'ordre supérieur, comme établi dans le prochain lemme.

Lemme 10.3.4. *Soit U une variable aléatoire positive. Alors, pour tout $n \geq 1$, on a :*

$$\mathbb{E}[U^n] = n \int_0^{+\infty} x^{n-1}\,\mathrm{P}\{U > x\}\,dx. \qquad (10.3.4)$$

Démonstration. En effet, d'après le théorème précédemment cité, on a : $\mathbb{E}[U^n] = \int_0^{+\infty} \mathrm{P}\{U^n > u\}\,du = \int_0^{+\infty} \mathrm{P}\{U > u^{1/n}\}\,du$; d'où la formule (10.3.4) en faisant le changement de variables $u^{1/n} = x$. □

Corollaire 10.3.5. *Soit (X_n) $(n \geq 0)$ une martingale telle que pour tout $n \geq 0$, on ait $\mathbb{E}[X_n^2] < +\infty$. Alors*

$$\mathbb{E}[(\max_{0 \le k \le n} |X_k|)^2] \le 4\,\mathbb{E}[X_n^2]. \tag{10.3.5}$$

Démonstration. On pose $A_n := \max_{0\le k\le n} |X_k|$, qui est donc une variable aléatoire *positive*. D'après le Corollaire 10.3.4, pour $n = 2$ et $U = A_n$, il vient : $\mathbb{E}[A_n^2] = 2\int_0^{+\infty} y\,\mathrm{P}\{A_n > y\}\,dy$. Comme, d'après le Corollaire 10.3.2, on a

$$\forall y > 0 \quad y\,\mathrm{P}\{A_n > y\} \le \mathbb{E}[\,|X_n|\,I_{\{A_n\,>\,y\}}],$$

on en déduit :

$$\mathbb{E}[A_n^2] \le 2\int_0^{+\infty} \mathbb{E}[\,|X_n|\,I_{\{A_n\,>\,y\}}]\,dy.$$

Soit $\mathrm{P}_{\mathbf{X}}$ la loi de probabilité du vecteur $\mathbf{X} := (X_0, \dots, X_n)$. En posant $a_n := \max_{0\le k\le n} |x_k|$ et $\mathbf{x} := (x_0, \dots, x_n)$, on a :

$$\mathbb{E}[\,|X_n|\,I_{\{A_n\,>\,y\}}] = \int_{\mathbb{R}^{n+1}} |x_n|\,I_{\{a_n\,>\,y\}}(\mathbf{x})\,d\,\mathrm{P}_{\mathbf{X}}(\mathbf{x}).$$

Par suite, les fonctions à intégrer étant positives,

$$\begin{aligned}\mathbb{E}[A_n^2] &\le 2\int_0^{+\infty}\Big(\int_{\mathbb{R}^{n+1}} |x_n|\,I_{\{a_n\,>\,y\}}(\mathbf{x})\,d\,\mathrm{P}_{\mathbf{X}}(\mathbf{x})\Big)dy\\ &= 2\int_{\mathbb{R}^{n+1}} |x_n|\,d\,\mathrm{P}_{\mathbf{X}}(\mathbf{x})\int_0^{+\infty} I_{\{a_n\,>\,y\}}(\mathbf{x})\,dy;\end{aligned}$$

d'où, puisque $\displaystyle\int_0^{+\infty} I_{\{a_n\,>\,y\}}(\mathbf{x})\,dy = \int_0^{+\infty} I_{\{a_n\,>\,y\}}(y)\,dy = \int_0^{a_n} dy = a_n$, la majoration

$$\begin{aligned}\mathbb{E}[A_n^2] &\le 2\int_{\mathbb{R}^{n+1}} |x_n|\,a_n\,d\,\mathrm{P}_{\mathbf{X}}(\mathbf{x}) = 2\,\mathbb{E}[\,|X_n|\,A_n]\\ &\le 2\sqrt{\mathbb{E}[X_n^2]\,\mathbb{E}[A_n^2]},\end{aligned}$$

d'après l'inégalité de Schwarz. On en déduit donc la majoration (10.3.5). □

Donnons encore une variante de l'inégalité maximale pour les sous-martingales de signe quelconque.

Théorème 10.3.6. *Soit (X_n) $(n \geq 0)$ une sous-martingale. Alors*

$$\forall \lambda > 0 \quad \lambda\,\mathrm{P}\{\min_{0\le k\le n} X_k < -\lambda\} \le \mathbb{E}[X_n^+] - \mathbb{E}[X_0].$$

Démonstration. Soient A l'évènement $\{\min_{0\le k\le n} X_k < -\lambda\}$ et T le temps d'arrêt défini par :

$$T := \min\{k : 0 \le k \le n,\ X_k < -\lambda\}\,I_A + n\,I_{A^c}.$$

Appliquons à la sous-martingale (X_n) $(n \geq 0)$ le théorème d'arrêt pour le temps d'arrêt *borné* $T \wedge n$ (*cf.* Théorème 8.2.1).

On obtient : $\mathbb{E}[X_0] \le \mathbb{E}[X_{T\wedge n}] = \mathbb{E}[X_T] = \mathbb{E}[X_T\, I_A] + \mathbb{E}[X_n\, I_{A^c}] \le -\lambda\,\mathbb{E}[I_A] + \mathbb{E}[X_n\, I_{A^c}] = -\lambda\,\mathrm{P}(A) + \mathbb{E}[X_n\, I_{\{-\lambda\le X_n<0\}} I_{A^c}] + \mathbb{E}[X_n\, I_{\{X_n\ge 0\}} I_{A^c}] \le -\lambda\,\mathrm{P}(A) + \mathbb{E}[X_n^+]$. □

10.4 LES INÉGALITÉS MAXIMALES POUR LES SURMARTINGALES

Nous donnons deux énoncés, un pour les surmartingales positives et un autre pour les surmartingales de signe quelconque.

Théorème 10.4.1. *Soit* (X_n) $(n \ge 0)$ *une surmartingale positive. Alors*

$$\forall\,\lambda > 0 \quad \lambda\,\mathrm{P}\{\max_{0\le k\le n} X_k > \lambda\} \le \mathbb{E}[X_0]. \tag{10.4.1}$$

Démonstration. Soient A l'évènement $\{\max_{0\le k\le n} X_k > \lambda\}$ et T le temps d'arrêt défini par :

$$T := \min\{k : 0 \le k \le n,\, X_k > \lambda\}\, I_A + n\, I_{A^c}.$$

Appliquons à la surmartingale (X_n) $(n \ge 0)$ le théorème d'arrêt pour le temps d'arrêt *borné* $T \wedge n$. On obtient : $\mathbb{E}[X_0] \ge \mathbb{E}[X_{T\wedge n}] = \mathbb{E}[X_T] = \mathbb{E}[X_T\, I_A] + \mathbb{E}[X_n\, I_{A^c}]$. Puisque les X_n sont *positifs*, on en déduit : $\mathbb{E}[X_0] \ge \mathbb{E}[X_T\, I_A] \ge \lambda\,\mathrm{P}(A)$. □

Remarque. L'inégalité (10.4.1) est uniforme en n. En faisant tendre n vers l'infini, il vient :

$$\forall\,\lambda > 0 \quad \lambda\,\mathrm{P}\{\max_{k\ge 0} X_k > \lambda\} \le \mathbb{E}[X_0].$$

Théorème 10.4.2. *Soit* (X_n) $(n \ge 0)$ *une surmartingale de signe quelconque. Alors*

$$\forall\,\lambda > 0 \quad \lambda\,\mathrm{P}\{\min_{0\le k\le n} X_k < -\lambda\} \le \mathbb{E}[X_n^+] - \mathbb{E}[X_n]. \tag{10.4.2}$$

Démonstration. Soient A l'évènement $\{\min_{0\le k\le n} X_k < -\lambda\}$ et T le temps d'arrêt défini par :

$$T := \min\{k : 0 \le k \le n,\, X_k < -\lambda\}\, I_A + n\, I_{A^c}.$$

Appliquons à la surmartingale (X_n) $(n \ge 0)$ le théorème d'arrêt pour le temps d'arrêt *borné* $T \wedge n$. On obtient, en utilisant l'inégalité $\mathbb{E}[X_0] \ge \mathbb{E}[X_{T\wedge n}] \ge \mathbb{E}[X_n]$, démontrée dans l'Exercice 8.1 : $\mathbb{E}[X_n] \le \mathbb{E}[X_{T\wedge n}] = \mathbb{E}[X_T] = \mathbb{E}[X_T\, I_A] + \mathbb{E}[X_n\, I_{A^c}] \le -\lambda\,\mathrm{P}(A) + \mathbb{E}[X_n^+]$. □

10.5 INÉGALITÉS DE NORMES

Pour tout réel $p \geq 1$ et toute variable aléatoire U, on pose

$$\|U\|_p := \Big(\int |U|^p \, d\,\mathrm{P}\Big)^{1/p} = (\,\mathbb{E}[\,|U|^p]\,)^{1/p}.$$

Soient U, V deux variables aléatoires définies sur le même espace. On a utilisé précédemment l'*inégalité de Schwarz* :

$$(\,\mathbb{E}[\,|U\,V|\,]\,)^2 \leq \mathbb{E}[U^2]\,\mathbb{E}[V^2]. \tag{10.5.1}$$

soit encore

$$(\|\,U\,V\,\|_1)^2 \leq \|\,U\,\|_2\,\|\,V\,\|_2. \tag{10.5.2}$$

On utilise aussi plus loin l'*inégalité de Hölder* sous la forme suivante : soient p, q deux nombres réels strictement positifs tels que $\dfrac{1}{p} + \dfrac{1}{q} = 1$. Alors

$$\|\,U\,V\,\|_1 \leq \|\,U\,\|_p\,\|\,V\,\|_q; \tag{10.5.3}$$

soit encore

$$\mathbb{E}[\,|U\,V|\,] \leq (\,\mathbb{E}[\,|U|^p]\,)^{1/p}\,(\,\mathbb{E}[\,|V|^q]\,)^{1/q}. \tag{10.5.4}$$

Soit (X_n) $(n \geq 0)$ une sous-martingale *positive*. Posons $U := \max\limits_{0 \leq k \leq n} X_k$. D'après le Théorème 10.3.1, on a les relations :

$$\forall\, y > 0 \quad y\,\mathrm{P}\{U > y\} \leq \mathbb{E}[X_n\,I_{\{U>y\}}]. \tag{10.5.5}$$

Soit p un nombre réel tel que $p > 1$ Comme, d'après le Lemme 10.3.4,

$$\mathbb{E}[U^p] = \int_0^{+\infty} py^{p-1}\,\mathrm{P}\{U > y\}\,dy,$$

on en déduit :

$$\mathbb{E}[U^p] \leq \int_0^{+\infty} py^{p-2}\,\mathbb{E}[X_n\,I_{\{U>y\}}]\,dy.$$

Posons $u := \max\limits_{0 \leq k \leq n} x_k$ et $\mathbf{x} := (x_0, \ldots, x_n)$, puis notons $\mathrm{P}_{\mathbf{X}}$ la loi de probabilité du vecteur $\mathbf{X} := (X_0, \ldots, X_n)$. Alors,

$$\mathbb{E}[X_n\,I_{\{U>y\}}] = \int_{\mathbb{R}_+^{n+1}} x_n\,I_{\{u>y\}}(\mathbf{x})\,d\,\mathrm{P}_{\mathbf{X}}(\mathbf{x}),$$

d'où

$$\mathbb{E}[U^p] \leq \int_{\mathbb{R}_+} py^{p-2}\,dy \int_{\mathbb{R}_+^{n+1}} x_n\,I_{\{u>y\}}(\mathbf{x})\,d\,\mathrm{P}_{\mathbf{X}}(\mathbf{x}).$$

Les fonctions à intégrer étant positives, on a, par le théorème de Fubini,

$$\begin{aligned}\mathbb{E}[U^p] &\leq \int_{\mathbb{R}_+^{n+1}} x_n\,d\,\mathrm{P}_{\mathbf{X}}(\mathbf{x}) \int_{\mathbb{R}_+} py^{p-2}\,I_{\{u>y\}}(\mathbf{x})\,dy \\ &= \int_{\mathbb{R}_+^{n+1}} x_n\,d\,\mathrm{P}_{\mathbf{X}}(\mathbf{x}) \int_0^u py^{p-2}\,dy;\end{aligned}$$

soit encore

$$\mathbb{E}[U^p] = \frac{p}{p-1}\int_{\mathbb{R}_+^{n+1}} x_n\, u^{p-1}\, d\,\mathrm{P}_{\mathbf{X}}(\mathbf{x});$$

d'où

$$\mathbb{E}[U^p] \leq \frac{p}{p-1}\,\mathbb{E}[X_n\, U^{p-1}]. \tag{10.5.6}$$

Prenons maintenant un nombre $q > 1$ tel que $\dfrac{1}{p} + \dfrac{1}{q} = 1$ et appliquons l'inégalité de Hölder (10.5.4) au membre de droite de (10.5.6). On obtient :

$$\mathbb{E}[\,|X_n\, U^{p-1}|\,] \leq (\,\mathbb{E}[\,X_n^p]\,)^{1/p}\,(\,\mathbb{E}[\,U^{(p-1)q}]\,)^{1/q}. \tag{10.5.7}$$

Comme $q = p/(p-1)$, on déduit de (10.5.6) et (10.5.7), l'inégalité :

$$\mathbb{E}[U^p] \leq q\,(\,\mathbb{E}[\,X_n^p]\,)^{1/p}\,(\,\mathbb{E}[\,U^p]\,)^{1/q}.$$

Si $\mathbb{E}[\,U^p]$ est fini et non nul, on en déduit l'inégalité

$$(\mathbb{E}[U^p])^{1/p} \leq q\,(\,\mathbb{E}[U^p]\,)^{1/q},$$

soit

$$\|U\|_p \leq q\,\|X_n\|_p. \tag{10.5.8}$$

Si $\mathbb{E}[U^p] = 0$, l'inégalité (10.5.8) est triviale. Si $\mathbb{E}[U^p]$ n'est pas fini, on considère la variable aléatoire bornée $U \wedge c$, de manière à obtenir $\|U \wedge c\|_p \leq q\,\|X_n\|_p$. On fait ensuite tendre c vers l'infini, pour obtenir (10.5.8) de nouveau. Récrivons le résultat obtenu sous forme de théorème.

Théorème 10.5.1. *Soient (X_n) $(n \geq 0)$ une sous-martingale positive et p un nombre satisfaisant $p > 1$. Alors*

$$\|\max_{0\leq k\leq n} X_k\|_p \leq \frac{p}{p-1}\,\|X_n\|_p. \tag{10.5.9}$$

Remarque. Si (X_n) $(n \geq 0)$ est une martingale, alors $(\,|X_n|\,)$ $(n \geq 0)$ est une sous-martingale positive et le précédent résultat s'applique. On a donc :

$$\|\max_{0\leq k\leq n} |X_k|\,\|_p \leq \frac{p}{p-1}\,\|\,|X_n|\,\|_p,$$

ou encore

$$\mathbb{E}[\max_{0\leq k\leq n} |X_k|^p\,] \leq \Big(\frac{p}{p-1}\Big)^p\,\mathbb{E}[\,|X_n|^p\,].$$

Corollaire 10.5.2. *Soient (X_n) $(n \geq 0)$ une sous-martingale positive et p un nombre satisfaisant $p > 1$. Alors*

$$\|\max_{0\leq k\leq n} X_k\|_1 \leq \frac{p}{p-1}\,\|X_n\|_p. \tag{10.5.10}$$

Si (X_n) $(n \geq 0)$ est une martingale, alors

$$\|\max_{0\leq k\leq n} |X_k|\,\|_1 \leq \frac{p}{p-1}\,\|\,|X_n|\,\|_p. \tag{10.5.11}$$

Démonstration. Si X est une variable aléatoire positive, l'application $p \mapsto \|X\|_p$ est *croissante*. Pour tout $p > 1$, on en déduit $\|X\|_1 \leq \|X\|_p$. Enfin, si (X_n) $(n \geq 0)$ est une martingale, on sait que $(|X_n|)$ $(n \geq 0)$ est une sous-martingale positive. □

Donnons un dernier résultat pour la norme $\|\cdots\|_1$. On pose :

$$\operatorname{Log}^+ x := \begin{cases} \operatorname{Log} x, & \text{si } x \geq 1\,; \\ 0, & \text{si } x < 1. \end{cases}$$

On observe (*cf.* Exercice 10.6) que pour $a \geq 0$ et $b > 0$, on a :

$$a \operatorname{Log} b \leq a \operatorname{Log}^+ a + \frac{b}{e}. \tag{10.5.12}$$

Théorème 10.5.3. *Si (X_n) $(n \geq 0)$ est une sous-martingale positive, alors*

$$\|\max_{0 \leq k \leq n} X_k\|_1 \leq \frac{e}{e-1}\,(1 + \mathbb{E}[X_n \operatorname{Log}^+ X_n\,]). \tag{10.5.13}$$

Si (X_n) $(n \geq 0)$ est une martingale, alors

$$\|\max_{0 \leq k \leq n} |X_k|\,\|_1 \leq \frac{e}{e-1}\,(1 + \mathbb{E}[\,|X_n| \operatorname{Log}^+ |X_n|\,]). \tag{10.5.14}$$

Démonstration. L'inégalité (10.5.13) entraîne (10.5.14) en utilisant l'argument désormais classique ! Pour (10.5.13), posons $U := \max_{0 \leq k \leq n} X_k$. Alors $\mathbb{E}[U] = \int_0^{+\infty} \mathrm{P}\{U > y\}\,dy = A + B$, avec $A := \int_0^1 \mathrm{P}\{U > y\}\,dy \leq 1$ et $B := \int_1^{+\infty} \mathrm{P}\{U > y\}\,dy$. Or, d'après (10.3.1), pour tout $y > 0$, on a :

$$y\,\mathrm{P}\{U > y\} \leq \mathbb{E}[X_n\, I_{\{U > y\}}].$$

D'où $B \leq \int_1^{+\infty} (1/y)\,\mathbb{E}[X_n\, I_{\{U>y\}}]\,dy$. Comme

$$\mathbb{E}[X_n\, I_{\{U>y\}}] = \int_{\mathbb{R}_+^{n+1}} x_n\, I_{\{u>y\}}\, d\,\mathrm{P}_{\mathbf{X}}(x_0, \ldots, x_n),$$

où $u := \max\limits_{0 \leq k \leq n} x_k$ et où $\mathrm{P}_{\mathbf{X}}$ est la loi de probabilité du vecteur $\mathbf{X} = (X_0, \ldots, X_n)$, on obtient, par le théorème de Fubini,

$$\begin{aligned} B &\leq \int_1^{+\infty} \frac{1}{y}\, dy \int_{\mathbb{R}_+^{n+1}} x_n\, I_{\{u>y\}}\, d\,\mathrm{P}_{\mathbf{X}}(x_0, \ldots, x_n) \\ &= \int_{\mathbb{R}_+^{n+1}} x_n\, d\,\mathrm{P}_{\mathbf{X}}(x_0, \ldots, x_n) \int_1^{+\infty} \frac{1}{y}\, I_{\{u>y\}}\, dy \\ &= \int_{\mathbb{R}_+^{n+1}} x_n\, I_{\{u \geq 1\}} \operatorname{Log} u\, d\,\mathrm{P}_{\mathbf{X}}(x_0, \ldots, x_n), \end{aligned}$$

soit, d'après (10.5.12),

$$B \leq \int_{\mathbb{R}_+^{n+1}} \Bigl(x_n \operatorname{Log}^+ x_n + \frac{u}{e}\Bigr)\, I_{\{u \geq 1\}}\, d\,\mathrm{P}_{\mathbf{X}}(x_0, \ldots, x_n).$$

On en tire

$$B \leq \mathbb{E}[X_n \text{ Log}^+ X_n] + \frac{1}{e}\int_{\mathbb{R}_+^{n+1}} u\, I_{\{u\geq 1\}}\, d\,\mathrm{P}_{\mathbf{X}}(x_0,\ldots,x_n)$$
$$\leq \mathbb{E}[X_n \text{ Log}^+ X_n] + \frac{1}{e}\,\mathbb{E}[U].$$

D'où

$$\mathbb{E}[U] = A + B$$
$$\leq 1 + \mathbb{E}[X_n \text{ Log}^+ X_n] + \frac{1}{e}\,\mathbb{E}[U],$$

soit

$$\Big(1-\frac{1}{e}\Big)\mathbb{E}[U] \leq 1 + \mathbb{E}[X_n \text{ Log}^+ X_n]. \quad \square$$

10.6 COMPLÉMENTS ET EXERCICES

10.1 Soit (X_n) $(n \geq 0)$ une martingale telle que pour un certain $p \geq 1$ et tout $n \geq 0$ on ait $\mathbb{E}[\,|X_n|^p\,] < +\infty$. Alors

$$\forall \varepsilon > 0 \quad \varepsilon^p\,\mathrm{P}\{\max_{0\leq k\leq n} |X_k| > \varepsilon\} \leq \mathbb{E}[\,|X_n|^p\,].$$

10.2 Soit (X_n) $(n \geq 0)$ une martingale telle que pour tout $n \geq 0$ on ait $\mathbb{E}[X_n] = 0$ et $\mathbb{E}[X_n^2] < +\infty$. Alors

$$\mathrm{P}\{\max_{0\leq k\leq n} X_k > \lambda\} \leq \frac{\mathbb{E}[X_n^2]}{\mathbb{E}[X_n^2] + \lambda^2}.$$

10.3 Suites de moyennes expérimentales
Soit (X_n) $(n \geq 1)$ une suite de variables aléatoires, indépendantes, identiquement distribuées, centrées, telles que $\mathbb{E}[\,|X_n|\,] < +\infty$ pour tout $n \geq 0$. Pour $n \geq 1$, on pose $S_n := X_1 + \cdots + X_n$, $Y_n := S_n/n$ et on note H_n la suite $H_n := (S_n, S_{n+1}, \ldots)$.

(a) La suite (Y_n) $(n \geq 1)$ est une martingale renversée par rapport à la suite (H_n) $(n \geq 1)$, c'est-à-dire la suite $(\ldots, Y_n, Y_{n-1}, \ldots, Y_1)$ est une martingale par rapport à la suite $(\ldots, H_n, H_{n-1}, \ldots, H_1)$. Autrement dit, on a : $\mathbb{E}[Y_n \mid H_{n+1}] = Y_{n+1}$.

(b) Pour tout $n \geq 1$ et tout $\lambda > 0$, on a : $\lambda\,\mathrm{P}\{\max_{k\geq n} |Y_k| > \lambda\} \leq \mathbb{E}[\,|Y_n|\,]$.

10.4 Partant de l'inégalité maximale (10.2.1), on démontre *directement* l'inégalité (10.5.10).

10.5 Soit (Y_n) $(n \geq 1)$ une suite de variables aléatoires indépendantes, identiquement disribuées, centrées et du second ordre. On pose : $Y_0 := 0$, $\mathbb{E}[Y_1^2] := \sigma^2 < +\infty$ et $S_n := Y_0 + Y_1 + \cdots + Y_n$ pour $n \geq 0$, de sorte que, pour tout $n \geq 0$, on a $\mathbb{E}[S_n] = 0$ et $\operatorname{Var} S_n = n\sigma^2$. On a la *généralisation* suivante de l'*inégalité de Kolmogorov* :

$$\forall \varepsilon > 0 \quad \varepsilon^2 \, \mathrm{P}\{\max_{0 \leq k \leq n} |S_k| > \varepsilon\} \leq \mathbb{E}[S_n^2 \, I_{\{\max_{0 \leq k \leq n} |S_n| > \varepsilon}].$$

10.6 On pose : $\operatorname{Log}^+ x := \begin{cases} \operatorname{Log} x, & \text{si } x \geq 1\,; \\ 0, & \text{si } x < 1. \end{cases}$

Pour $a \geq 0$ et $b > 0$, démontrer l'inégalité :

$$a \operatorname{Log} b \leq a \operatorname{Log}^+ a + \frac{b}{e}.$$

Chapitre 11

Théorème de convergence

Dans le théorème de convergence des martingales, on donne une condition suffisante pour qu'une martingale converge vers une variable aléatoire, qui soit intégrable. La technique de démonstration du théorème fait appel à un argument de comptage. Il s'agit de compter le nombre de fois qu'une trajectoire traverse une bande horizontale et de majorer ce nombre.

11.1 LE THÉORÈME DE CONVERGENCE

Théorème 11.1.1. *Soit (X_n) $(n \geq 0)$ une martingale bornée dans L^1, c'est-à-dire telle que, pour une certaine constante $c > 0$, on ait :*

$$\sup_{n\geq 0} \mathbb{E}[\,|X_n|\,] \leq c. \tag{11.1.1}$$

Alors la suite (X_n) $(n \geq 0)$ converge presque sûrement vers une variable aléatoire intégrable X_∞.

La démonstration utilise un lemme dû à Doob sur la majoration de la moyenne des traversées d'une bande horizontale par la suite (X_n). Pour définir cette notion, nous supposons donnés deux nombres a, b tels que $a < b$ et introduisons la suite $S_1 < T_1 < S_2 < T_2 < \cdots$ des temps d'arrêt suivante :

$$S_1 := \inf\{n \geq 0 : X_n \leq a\}, \quad T_1 := \inf\{n > S_1 : X_n \geq b\},$$
$$S_2 := \inf\{n > T_1 : X_n \leq a\}, \quad T_2 := \inf\{n > S_2 : X_n \geq b\},$$

et ainsi de suite. Si l'une des bornes inférieures n'existe pas, on donne la valeur $+\infty$ au temps d'arrêt correspondant, ainsi qu'aux suivants. Dans la

Figure 11.1, on a représenté une trajectoire $n \mapsto X_n$ par des croix "×". Les temps d'arrêt $S_1, T_1, S_2, \ldots$ sont représentés par des gros points "•".

La variable aléatoire

$$M := \sum_{k \geq 1} I_{\{T_k < +\infty\}} \tag{11.1.2}$$

est le nombre total de *traversées de* $[a, b]$, *en montant,* effectuées par la trajectoire $n \mapsto X_n$.

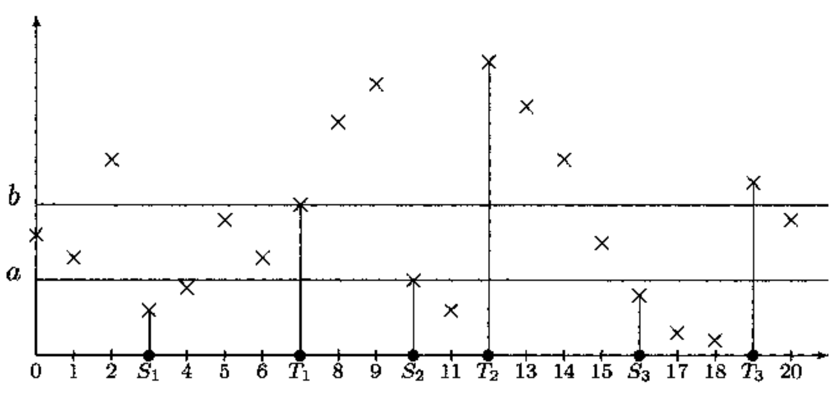


FIGURE 11.1

Lemme 11.1.2 (Inégalité de Dubins). *Pour tout $k \geq 1$ et tout $n \geq 1$, on a :*

$$(b - a)\,\mathrm{P}\{T_k < n\} \leq \mathbb{E}[(a - X_n)\, I_{\{S_k \leq n < T_k\}}]. \tag{11.1.3}$$

Démonstration. L'entier n étant fixé, posons

$$D_k := X_{T_k \wedge n} - X_{S_k \wedge n}.$$

D'après le théorème d'arrêt appliqué aux temps d'arrêt *bornés* $T_k \wedge n$ et $S_k \wedge n$, on voit que : $\mathbb{E}[D_k] = \mathbb{E}[X_1] - \mathbb{E}[X_1] = 0$. D'autre part, on a les implications

$$\{n < S_k\} \Rightarrow \{D_k = 0\}; \quad \{T_k \leq n\} \Rightarrow \{D_k \geq b - a\};$$
$$\{S_k \leq n < T_k\} \Rightarrow \{D_k = X_n - X_{S_k}\} \Rightarrow \{D_k \geq X_n - a\};$$

qu'on peut résumer en écrivant :

$$(b - a)\, I_{\{T_k \leq n\}} + (X_n - a)\, I_{\{S_k \leq n < T_k\}} \leq D_k.$$

En prenant l'espérance mathématique des deux membres, on obtient :

$$(b - a)\,\mathrm{P}\{T_k \leq n\} + \mathbb{E}[(X_n - a)\, I_{\{S_k \leq n < T_k\}}] \leq \mathbb{E}[D_k] = 0. \quad \square$$

Lemme 11.1.3. (Lemme de Doob). *Conservant les notations* (11.1.1) *et* (11.1.2), *on a la majoration :*

$$\mathbb{E}[M] \leq \frac{|a| + c}{b - a}. \tag{11.1.4}$$

Démonstration. L'inégalité de Dubins (11.1.3) permet d'écrire

$$(b-a)\sum_{k\geq 1}\mathrm{P}\{T_k<n\}\leq \sum_{k\geq 1}\mathbb{E}[(a-X_n)\,I_{\{S_k\leq n<T_k\}}].$$

Or, les évènements $\{S_k\leq n<T_k\}$ $(k\geq 1)$ sont disjoints deux à deux; notons A leur réunion. Comme $\{S_k\leq n<T_k\}=\emptyset$ dès que $k\geq n$, on peut écrire :

$$\begin{aligned}\sum_{k\geq 1}\mathbb{E}[(a-X_n)\,I_{\{S_k\leq n<T_k\}}]&=\sum_{1\leq k\leq n}\mathbb{E}[(a-X_n)\,I_{\{S_k\leq n<T_k\}}]\\&=\mathbb{E}[(a-X_n)\sum_{1\leq k\leq n}I_{\{S_k\leq n<T_k\}}]\\&=\mathbb{E}[(a-X_n)\sum_{k\geq 1}I_{\{S_k\leq n<T_k\}}]\\&=\mathbb{E}[(a-X_n)I_A];\end{aligned}$$

d'où

$$(b-a)\sum_{k\geq 1}\mathrm{P}\{T_k<n\}\leq \mathbb{E}[(a-X_n)I_A].$$

Maintenant, on a les majorations : $\mathbb{E}[(a-X_n)\,I_A]\leq \mathbb{E}[(a-X_n)^+\,I_A]\leq \mathbb{E}[(a-X_n)^+]\leq |a|+c$, d'où, pour tout n la majoration $\sum_{k\geq 1}\mathrm{P}\{T_k<n\}\leq A:=(|a|+c)/(b-a)$. D'autre part, $\mathbb{E}[M]=\sum_{k\geq 1}\mathrm{P}\{T_k<+\infty\}$. Si l'on avait $\mathbb{E}[M]>A$, on aurait aussi $\sum_{1\leq k\leq K}\mathrm{P}\{T_k<+\infty\}>A$ pour un certain $K\geq 1$. Comme, pour tout $k\geq 1$, on a $\mathrm{P}\{T_k<+\infty\}=\lim_n\mathrm{P}\{T_k<n\}$, on aurait $\sum_{1\leq k\leq K}\mathrm{P}\{T_k<n\}>A$ pour n assez grand, a fortiori $\sum_{k\geq 1}\mathrm{P}\{T_k<n\}>A$, ce qui contredit la majoration prouvée pout tout n. □

Lemme 11.1.4. *La suite (X_n) $(n\geq 0)$ est presque sûrement convergente.*

Démonstration. Pour indiquer que la variable aléatoire M dépend du couple (a,b), on pose : $M:=M_{a,b}$. C'est une variable aléatoire positive et l'inégalité (11.1.4) montre qu'elle est intégrable. Elle est donc presque sûrement finie, ou encore l'évènement $\{M_{a,b}=+\infty\}$ est négligeable. La réunion *dénombrable*

$$\bigcup_{a,b\in\mathbb{Q},\,a<b}\{M_{a,b}=+\infty\}$$

est aussi négligeable. Or, l'évènement $\{\liminf_n X_n<a<b<\limsup_n X_n\}$ entraîne qu'il y a une infinité d'indices n tels que $\{X_n<a\}$ se produise et qu'il y a aussi une infinité d'indices n tels que $\{X_n>b\}$ se produise également; ce qui implique alors que l'évènement $\{M_{a,b}=+\infty\}$ se réalise. On a donc

$$\{\liminf_n X_n<a<b<\limsup_n X_n\}\subset\{M_{a,b}=+\infty\}.$$

D'où

$$\{\liminf_n X_n < \limsup_n X_n\} = \bigcup_{a,b\in\mathbb{Q},\, a<b} \{\liminf_n X_n < a < b < \limsup_n X_n\}$$
$$\subset \bigcup_{a,b\in\mathbb{Q},\, a<b} \{M_{a,b} = +\infty\}.$$

L'évènement $\{\liminf_n X_n < \limsup_n X_n\}$ est donc négligeable, ce qui signifie que la suite (X_n) $(n \geq 0)$ converge presque sûrement. □

Fin de la démonstration du Théorème 11.1.1. Il reste à démontrer que la suite (X_n) $(n \geq 0)$ converge presque sûrement vers une variable aléatoire *intégrable*. Cette propriété résulte du Lemme de Fatou (*cf.* [FF1], chap. 10, Corollaire du Théorème 9.2).

En effet, d'après le Lemme 11.1.4, la suite (X_n) converge presque sûrement, disons, vers une variable aléatoire X_∞, d'où $|X_n| \xrightarrow{p.s.} |X_\infty|$. Ensuite, pour tout $n \geq 0$, on a $\mathbb{E}[\,|X_n|\,] \leq c$. Les deux conditions du Lemme de Fatou sont remplies et on peut conclure que l'on a aussi $\mathbb{E}[\,|X_\infty|\,] \leq c$. □

Remarque 1. La démonstration du Théorème 11.1.1 reste valable, si X est une surmartingale. En effet, on a alors $\mathbb{E}[D_n] \leq 0$ et les inégalités restent toutes dans le même sens. Pour vérifier l'inégalité $\mathbb{E}[D_n] \leq 0$, on utilise le fait que, pour une surmartingale admettant deux temps d'arrêt S, T tels que $S \leq T$, alors $\mathbb{E}[X_S] \geq \mathbb{E}[X_T]$. Intuitivement, on comprend que, pour une surmartingale, on puisse majorer le nombre de traversées en montant, puisqu'une surmartingale a tendance à décroître.

Remarque 2. Si X est une *sous*-martingale, alors $-X$ est une surmartingale. À toute traversée en montant de $[a, b]$ par la surmartingale $-X$ correspond une traversée en descendant de $[-b, -a]$ par la sous-martingale X. Le Théorème 11.1.1 appliqué à la surmartingale $-X$ fournit alors une majoration pour le nombre de traversées en descendant de $[-b, -a]$ par la sous-martingale X. Le Théorème 11.1.1 reste donc encore valable aussi pour une sous-martingale.

Corollaire 11.1.5. *Une martingale de signe constant (resp. une surmartingale positive, resp. une sous-martingale négative ou nulle) est presque sûrement convergente.*

Démonstration. Démontrons le corollaire pour une martingale *positive*. On a, pour tout $n \geq 0$ les relations $\mathbb{E}[\,|X_n|\,] = \mathbb{E}[X_n] = \mathbb{E}[X_0] < +\infty$ et le Théorème 1.1 s'applique. Pour une surmartingale positive, $\mathbb{E}[\,|X_n|\,] = \mathbb{E}[X_n] \leq \mathbb{E}[X_0] < +\infty$ et pour une sous-martingale $X \leq 0$, il vient : $\mathbb{E}[\,|X_n|\,] = -\mathbb{E}[X_n] \leq -\mathbb{E}[X_0] < +\infty$. □

11.2 MARTINGALES UNIFORMÉMENT INTÉGRABLES

Rappelons qu'une suite de variables aléatoires (X_n) $(n \geq 0)$ est *uniformément intégrable* (*cf.* chap. 8, § 5), si

$$\lim_{c \to 0} \sup_n \mathbb{E}[I_{\{|X_n|>c\}}|X_n|] = 0. \tag{11.2.1}$$

Donnons, d'abord, quelques conditions suffisantes pour qu'une suite de variables aléatoires soit uniformément intégrable.

Proposition 11.2.1. *Soient (X_n) $(n \geq 0)$ une suite de variables aléatoires et Z une variable aléatoire positive, intégrable, telles que pour tout $n \geq 0$ on ait $|X_n| \leq Z$. Alors la suite (X_n) $(n \geq 0)$ est uniformément intégrable.*

Démonstration. Pour tout $c > 0$ et tout $n \geq 0$, on a

$$\mathbb{E}[I_{\{|X_n|>c\}}|X_n|] \leq \mathbb{E}[Z\, I_{\{Z>c\}}];$$

donc, pour tout $c > 0$

$$\sup_n \mathbb{E}[I_{\{|X_n|>c\}}|X_n|] \leq \mathbb{E}[Z\, I_{\{Z>c\}}]. \tag{11.2.2}$$

Or, puisque Z est positive, on a : $\lim_n \mathbb{E}[Z\, I_{\{Z \leq n\}}] = \mathbb{E}[\lim_n Z\, I_{\{Z \leq n\}}] = \mathbb{E}[Z] < +\infty$. Donc, pour c assez grand, le membre de droite de (11.2.2) est arbitrairement petit. □

Proposition 11.2.2. *Supposons que pour un certain $\alpha > 0$ et pour tout $n \geq 0$, il existe une constante finie K telle que $\mathbb{E}[\,|X_n|^{1+\alpha}] \leq K$. Alors, la suite (X_n) $(n \geq 0)$ est uniformément intégrable.*

Démonstration. On fait appel ici à l'inégalité de Hölder sous la forme :

$$\mathbb{E}[\,|XY|\,] \leq (\mathbb{E}[\,|X|^r\,])^{1/r}(\mathbb{E}[\,|Y|^s\,])^{1/s},$$

où r et s sont deux nombres supérieurs à 1 tels que $(1/r) + (1/s) = 1$. Prenons $r := 1 + \alpha$. Alors, l'inégalité de Hölder donne :

$$\begin{aligned}\mathbb{E}[I_{\{|X_n|>c\}}|X_n|] &\leq (\mathbb{E}[\,|X_n|^r\,])^{1/r}(\mathbb{E}[\,I^s_{\{|X_n|>c\}}\,])^{1/s}\\ &\leq K^{1/r}\,(\mathrm{P}\{|X_n| > c\})^{1/s}\\ &\leq K^{1/r}\left(\frac{\mathbb{E}[\,|X_n|^r\,]}{c^r}\right)^{1/s} \quad \text{[par l'inégalité de Tchebychev]}\\ &\leq K^{1/r}\left(\frac{K}{c^r}\right)^{1/s} = \frac{K}{c^{r/s}}.\end{aligned}$$

Comme $r/s > 0$, la dernière fraction est arbitrairement petite pour c assez grand. □

La prochaine proposition est la réciproque de la précédente.

Proposition 11.2.3. *Si (X_n) $(n \geq 0)$ est une suite de variables aléatoires uniformément intégrable, alors $\sup_n \mathbb{E}[\,|X_n|\,] = K < +\infty$.*

Démonstration. En effet, pour tout $c > 0$ et tout $n \geq 0$, on a

$$\begin{aligned}\mathbb{E}[\,|X_n|\,] &= \mathbb{E}[I_{\{|X_n|\leq c\}}|X_n|\,] + \mathbb{E}[\,I_{\{|X_n|>c\}}|X_n|\,]\\ &\leq c + \mathbb{E}[\,I_{\{|X_n|>c\}}|X_n|\,],\end{aligned}$$

d'où, pour tout $c > 0$,

$$\sup_n \mathbb{E}[\,|X_n|\,] \leq c + \sup_n \mathbb{E}[\,I_{\{|X_n|>c\}}|X_n|\,].$$

Comme $\lim\limits_{c\to 0} \sup_n \mathbb{E}[\,I_{\{|X_n|>c\}}|X_n|\,] = 0$, on peut trouver un nombre $C > 0$ tel que $\sup_n \mathbb{E}[\,I_{\{|X_n|>C\}}|X_n|\,] \leq 1$, d'où $\sup_n \mathbb{E}[\,|X_n|\,] \leq C + 1$. □

La prochaine proposition est donnée sans démonstration.

Proposition 11.2.4. *Étant donnée une suite de variables aléatoires (X_n) $(n \geq 0)$, les deux propriétés suivantes sont équivalentes :*

(a) *la suite (X_n) $(n \geq 0)$ est convergente en moyenne d'ordre* 1 *vers une variable aléatoire X intégrable ;*

(b) $X_n \xrightarrow{p} X$ *et la suite (X_n) $(n \geq 0)$ est uniformément intégrable.*

Il faut noter que la propriété d'*intégrabilité uniforme* est exactement la propriété qui fait défaut à la propriété de convergence en probabilité pour entraîner la convergence en moyenne d'ordre 1.

Le corollaire suivant est une conséquence directe des Propositions 11.2.1 et 11.2.4.

Corollaire 11.2.5. *Soit (X_n) $(n \geq 0)$ une suite de variables aléatoires telles que pour tout $n \geq 0$ on ait $|X_n| \leq Z$, où Z est une variable aléatoire intégrable. Alors la convergence en probabilité est équivalente à la convergence en moyenne d'ordre* 1. *Un cas particulier important est le cas où la suite (X_n) $(n \geq 0)$ est bornée.*

Théorème 11.2.6 (Théorème de convergence pour les martingales uniformément intégrables). *Soit (X_n) $(n \geq 0)$ une martingale uniformément intégrable. Alors il existe une variable aléatoire intégrable X_∞ telle que*

(a) $X_n \xrightarrow{p.s.} X_\infty$;

(b) $X_n \to X_\infty$ *(en moyenne d'ordre* 1*), i.e.* $\mathbb{E}[\,|X_n - X_\infty|\,] \to 0$.

En outre, pour tout $n \geq 0$, on a : $\mathbb{E}[X_\infty] = \mathbb{E}[X_n]$.

Démonstration. D'après la Proposition 11.2.3, l'hypothèse du Théorème 1.1 est vérifiée. On a donc $X_n \xrightarrow{p.s.} X_\infty$, d'où $X_n \xrightarrow{p} X_\infty$. D'après la Proposition 2.4, on a donc aussi $X_n \to X_\infty$ en moyenne d'ordre 1 et X_∞ est intégrable. Alors $|\mathbb{E}[X_n] - \mathbb{E}[X_\infty]\,| \leq \mathbb{E}[\,|X_n - X_\infty|\,] \to 0$. D'où $\mathbb{E}[X_n] \to \mathbb{E}[X_\infty]$. Or, la suite (X_n) $(n \geq 0)$ étant une martingale, la suite $(\mathbb{E}[X_n])$ $(n \geq 0)$ est constante. On a donc : $\mathbb{E}[X_\infty] = \mathbb{E}[X_n]$. □

Le prochain théorème indique le lien étroit qui lie intégrabilité uniforme et martingale de Doob. Auparavant, nous établissons, d'abord, un lemme de *continuité absolue*, qui est une conséquence du théorème de convergence monotone de Beppo Levi (*cf.* [FF1], chap. 10, Théorème 10.1).

Lemme 11.2.7. *Si Z est une variable aléatoire intégrable, alors, si* $\mathrm{P}(A)$ *tend vers* 0, *on a* $\mathbb{E}[I_A\,|Z|] \to 0$.

Démonstration. En effet, $\mathbb{E}[I_{\{|Z|<n\}}\,|Z|]$ tend en croissant vers $\mathbb{E}[\,|Z|\,]$, d'après le théorème de Beppo Levi. Si $\varepsilon > 0$ est donné, il existe N tel que $\mathbb{E}[\,|Z|\,] \le \mathbb{E}[I_{\{|Z|<N\}}\,|Z|\,] + \varepsilon$. Prenons A tel que $\mathrm{P}(A) \le \varepsilon/N$, ce qui implique que $\mathbb{E}[I_A I_{\{|Z|<N\}}\,|Z|] \le N\mathbb{E}[I_A] = N\mathrm{P}(A) \le \varepsilon$. Alors

$$\begin{aligned}\mathbb{E}[I_A\,|Z|\,] &= \mathbb{E}[I_A I_{\{|Z|<N\}}\,|Z|\,] + (\mathbb{E}[I_A\,|Z|\,] - \mathbb{E}[I_A I_{\{|Z|<N\}}\,|Z|\,]) \\ &\le \varepsilon + \mathbb{E}[\,|Z|\,] - \mathbb{E}[I_{\{|Z|<N\}}\,|Z|\,] < 2\,\varepsilon. \quad \square\end{aligned}$$

Théorème 11.2.8. *Soit* (X_n) $(n \ge 0)$ *une martingale par rapport à une suite* (Y_n) $(n \ge 0)$. *Alors les deux propriétés suivantes sont équivalentes :*

(a) (X_n) $(n \ge 0)$ *est une martingale de Doob, i.e. il existe une variable aléatoire intégrable Z telle que pour tout $n \ge 0$ on ait* $X_n = \mathbb{E}[Z \mid Y_0, \dots, Y_n]$ *;*

(b) (X_n) $(n \ge 0)$ *est uniformément intégrable.*

Démonstration. Supposons (a) satisfait et faisons l'hypothèse $Z \ge 0$, le cas général, que nous ne traitons pas, s'en déduisant en considérant $Z = Z^+ - Z^-$. Dans ce cas, les variables aléatoires X_n sont positives.

D'abord, appliquant les règles de l'espérance conditionnelle,

$$\begin{aligned}\mathbb{E}[\,I_{\{X_n>c\}}X_n\,] &= \mathbb{E}[\,I_{\{X_n>c\}}\mathbb{E}[Z \mid Y_0, \dots, Y_n]\,] \\ &= \mathbb{E}[\,\mathbb{E}[I_{\{X_n>c\}}\,Z \mid Y_0, \dots, Y_n]\,] \\ &= \mathbb{E}[I_{\{X_n>c\}}\,Z]. \end{aligned} \tag{11.2.3}$$

En particulier, en prenant, par exemple, $c = -1$, on obtient $\mathbb{E}[X_n] = \mathbb{E}[Z]$.

D'après l'inégalité de Tchebychev, pour tout $n \ge 0$, on a :

$$\mathrm{P}\{X_n > c\} \le \frac{1}{c}\,\mathbb{E}[X_n] = \frac{1}{c}\,\mathbb{E}[Z],$$

d'où

$$\sup_n \mathrm{P}\{X_n > c\} \le \frac{1}{c}\,\mathbb{E}[Z],$$

et donc

$$\lim_{c\to+\infty} \sup_n \mathrm{P}\{X_n > c\} = 0. \tag{11.2.4}$$

Donc, quel que soit $n \ge 0$, on a $\lim_{c\to+\infty} \mathrm{P}\{X_n > c\} = 0$, ce qui entraîne, d'après le Lemme 11.2.7, que $\lim_{c\to+\infty} \mathbb{E}[I_{\{X_n>c\}}\,Z] = 0$, ou encore, d'après (11.2.3), que $\lim_{c\to+\infty} \mathbb{E}[I_{\{X_n>c\}}\,X_n] = 0$. D'où $\lim_{c\to+\infty} \sup_n \mathbb{E}[I_{\{X_n>c\}}\,X_n] = 0$. La suite (X_n) $(n \ge 0)$ est donc uniformément intégrable.

Réciproquement, supposons cette dernière condition satisfaite. D'après le Théorème 11.2.6, il existe une variable aléatoire intégrable X_∞ telle que $\mathbb{E}[\,|X_n - X_\infty|\,] \to 0$. Nous nous proposons de démontrer que, pour tout $n \ge 0$, on a : $X_n = \mathbb{E}[X_\infty \mid Y_0, \dots, Y_n]$. Il suffit donc, d'après la Proposition 1.13.7, de démontrer que, pour tout entier $n \ge 0$ et tout évènement $A \in \mathfrak{T}(Y_0, \dots, Y_n)$, on a : $\mathbb{E}[I_A X_n] = \mathbb{E}[I_A X_\infty]$.

D'après la Proposition 7.2.4, pour n et $k \geq 0$, on a :

$$\mathbb{E}[X_{n+k} \mid Y_0, \ldots, Y_n] = \mathbb{E}[X_n \mid Y_0, \ldots, Y_n].$$

D'où

$$\begin{aligned}\mathbb{E}[I_A X_n] &= \mathbb{E}[\, I_A \mathbb{E}[X_n \mid Y_0, \ldots, Y_n]\,] = \mathbb{E}[\, I_A \mathbb{E}[X_{n+k} \mid Y_0, \ldots, Y_n]\,] \\ &= \mathbb{E}[\, \mathbb{E}[I_A X_{n+k} \mid Y_0, \ldots, Y_n]\,] = \mathbb{E}[I_A X_{n+k}].\end{aligned}$$

La suite de terme général $\mathbb{E}[I_A X_n]$ est donc constante. Comme $|\,\mathbb{E}[I_A X_n] - \mathbb{E}[I_A X_\infty]\,| \leq \mathbb{E}[\,|X_n - X_\infty|\,]$ et que $\lim_n \mathbb{E}[\,|X_n - X_\infty|\,] = 0$, la suite de terme général $\mathbb{E}[I_A X_n]$ tend vers $\mathbb{E}[I_A X_\infty]$. D'où $\mathbb{E}[I_A X_n] = \mathbb{E}[I_A X_\infty]$. □

11.3 COMPLÉMENTS ET EXERCICES

11.1 Une martingale de Doob est presque sûrement convergente
Une variable aléatoire Z vérifiant $\mathbb{E}[\,|Z|\,] < +\infty$, étant donnée, ainsi qu'une suite de variables aléatoires (Y_n) $(n \geq 0)$, toutes définies sur le même espace probabilisé que Z, on a posé $X_n := \mathbb{E}[Z \mid Y_0, \ldots, Y_n]$ pour tout $n \geq 0$ (*cf.* § 7.2) et démontré que la suite (X_n) $(n \geq 0)$ est une martingale, dite martingale de Doob. Alors la suite (X_n) converge presque sûrement vers une variable aléatoire X_∞.

11.2 Soient Z une variable aléatoire telle que $\mathbb{E}[\,|Z|\,] < +\infty$ et (X_n) $(n \geq 0)$ une martingale par rapport à une suite (Y_n) $(n \geq 0)$ telle que, pour tout $n \geq 0$, on ait $X_n \leq Z$. Alors, la suite (X_n) $(n \geq 0)$ converge presque sûrement. [Considérer la martingale de Doob $(\mathbb{E}[Z \mid Y_0, \ldots, Y_n])$ $(n \geq 0)$.]

11.3 Soit M le nombre de traversées en montant de l'intervalle $[a, b]$ d'une martingale (X_n) $(n \geq 0)$, comme défini en (11.1.2). Conservons les notations du paragraphe 1. Alors, $\text{Trav}_k := \{S_k < +\infty, T_k = +\infty\}$ est l'évènement “la trajectoire $n \mapsto X_n$ admet $(k-1)$ traversées en montant et la $k^{\text{ième}}$ est seulement amorcée”. On a l'inégalité

$$\forall\, k \geq 1 \quad (b-a)\, \mathrm{P}\{M \geq k\} \leq \mathbb{E}[(a - X_n)\, I_{\text{Trav}_k}\,].$$

[Appliquer l'inégalité de Dubins à la martingale arrêtée $X^{\mid n}$.]

11.4 Soit (Y_n) $(n \geq 1)$ une suite de variables aléatoires indépendantes, identiquement distribuées, à valeurs positives, du premier ordre, de loi non-dégénérée (à savoir, $\mathcal{L}(Y_1) \neq \varepsilon_1$) et vérifiant $\mathbb{E}[Y_1] = 1$. Pour chaque $n \geq 1$, on pose : $X_n := \prod_{k=1}^{n} Y_k$.

(a) La suite (X_n) $(n \geq 1)$ est une martingale positive par rapport à (Y_n) $(n \geq 1)$.

(b) D'après (a), la suite (X_n) $(n \geq 1)$ converge presque sûrement. De plus, $X_n \xrightarrow{p.s.} 0$ lorsque n tend vers l'infini. [Distinguer deux cas : $\mathrm{P}\{Y_1 = 0\} > 0$ et $\mathrm{P}\{Y_1 = 0\} = 0$.]

11.5 Modèle de Polyà

À la date 0, une urne contient une boule blanche et une boule noire. À la date 1, on prélève une boule, on note sa couleur, puis on la remet dans l'urne, avec une boule supplémentaire *de même couleur*. On itère cette opération aux dates 2, 3, … Soit N_n $(n \geq 0)$ le nombre de boules blanches dans l'urne après la $n^{\text{ième}}$ opération.

(a) La suite de terme général $X_n := \dfrac{N_n}{n+2}$ est une martingale par rapport à (N_n) $(n \geq 0)$.

(b) Le rapport du nombre de boules blanches au nombre de boules noires converge presque sûrement vers une limite.

Chapitre 12

Exemples d'applications

On se propose, dans ce chapitre, de donner cinq exemples d'applications de la théorie des chaînes de Markov et des martingales. Les *processus de branchement* font l'objet, en général, d'une étude spécifique. Il paraît intéressant ici de voir comment les résultats sur les chaînes de Markov et les martingales s'appliquent dans ce cadre. La *loi de Hardy-Weinberg*, qui est traitée ensuite, fournit une application inattendue en génétique. On verra également que le *battage des cartes* peut s'interpréter commodément dans le cadre des chaînes de Markov. Enfin, une extension du traditionnel *problème du scrutin*, ainsi que celui du *collectionneur de vignettes*, trouvent une solution élégante dans le cadre des martingales.

12.1 PROCESSUS DE BRANCHEMENT

On considère une particule qui donne naissance à un nombre aléatoire X de particules de même espèce. On pose $m := \mathbb{E}[X]$ et l'on suppose $0 < m < +\infty$. On fait l'hypothèse que chaque particule de la première génération donne naissance, à son tour, à un nombre aléatoire de particules de même espèce, ces nombres aléatoires étant indépendants et tous de même loi de probabilité que X. Les particules de la deuxième génération se reproduisent de nouveau et ainsi de suite. On désigne par Y_n $(n \geq 0)$ le nombre de particules de la $n^{\text{ième}}$ génération. On suppose $Y_0 := 1$. Autrement dit, la génération initiale est formée *d'un seul* individu.

Lemme 12.1.1. *On a :* $\mathbb{E}[Y_n] = m^n$ $(n \geq 0)$.

Démonstration. Notons $X^{(n)}(i)$ le nombre de particules produites par le $i^{\text{ième}}$ parent de la $n^{\text{ième}}$ génération. Les variables aléatoires $X^{(n)}(1)$, $X^{(n)}(2)$, ... sont *indépendantes, identiquement distribuées*, de loi commune $\mathcal{L}(X)$. Le nombre de particules de la $(n+1)^{\text{ième}}$ génération est alors donné par la somme aléatoire : $Y_{n+1} = X^{(n)}(1) + \cdots + X^{(n)}(Y_n)$. D'après l'identité de Wald, on a donc : $\mathbb{E}[Y_{n+1}] = \mathbb{E}[Y_n]\,\mathbb{E}[X] = \mathbb{E}[Y_n]\,m$ $(n \geq 0)$ et enfin $\mathbb{E}[Y_n] = m^n$. □

Lorsque n tend vers l'infini, le comportement asymptotique de $\mathbb{E}[Y_n]$ se déduit directement du Lemme 12.1.1. Suivant que $m > 1$ ou que $m < 1$, le nombre moyen $\mathbb{E}[Y_n]$ tend *exponentiellement* vers l'infini ou vers 0. Il reste constamment égal à 1 si $m = 1$.

Notons que si $m < 1$, on a $\sum_{n\geq 0} \mathbb{E}[Y_n] = \sum_{n\geq 0} m^n = 1/(1-m)$, une quantité qui représente l'espérance mathématique du nombre total d'individus issus de l'ancêtre commun de la génération 0 (cet ancêtre étant compris) et ce nombre est *fini*.

Lemme 12.1.2. *La suite (Y_n) $(n \geq 0)$ est une chaîne de Markov, homogène.*

Ce lemme est évident, car le nombre de particules de la $(n+1)^{\text{ième}}$ génération ne dépend que du nombre de particules de la $n^{\text{ième}}$ génération ; il est, de plus, indépendant de n.

Lemme 12.1.3. *On pose : $Z_n := Y_n/m^n$ $(n \geq 0)$. Le processus (Z_n) $(n \geq 0)$ est une martingale par rapport à (Y_n) $(n \geq 0)$.*

Démonstration. D'après le Lemme 12.1.1, on a : $\mathbb{E}[\,|Z_n|\,] = \mathbb{E}[Z_n] = 1 < +\infty$ pour tout $n \geq 0$. Il résulte, ensuite, du Lemme 12.1.2 que l'on a : $\mathbb{E}[Y_{n+1} \,|\, Y_0 = i_0, \ldots, Y_n = i_n] = \mathbb{E}[Y_{n+1} \,|\, Y_n = i_n] = \mathbb{E}[X^{(n)}(1) + \cdots + X^{(n)}(i_n)] = i_n\,\mathbb{E}[X] = i_n\,m$, ce qui implique $\mathbb{E}[Y_{n+1} \,|\, Y_0, \ldots, Y_n] = Y_n\,m$, c'est-à-dire, en introduisant la variable aléatoire Z_n, l'identité recherchée : $\mathbb{E}[Z_{n+1} \,|\, Y_0, \ldots, Y_n] = (1/m^{n+1})\mathbb{E}[Y_{n+1} \,|\, Y_0, \ldots, Y_n] = Y_n/m^n = Z_n$. □

Remarque. Le fait que (Z_n) $(n \geq 0)$ soit une martingale permet d'établir directement que l'on a $\mathbb{E}[Y_n] = m^n$, puisque $\mathbb{E}[Z_n] = \mathbb{E}[Z_0] = 1$.

Lemme 12.1.4. *La suite (Z_n) $(n \geq 0)$ converge presque sûrement vers une variable aléatoire limite, positive, intégrable Z_∞. Si $m > 1$, la suite (Y_n) $(n \geq 0)$ croît exponentiellement lorsque n croît sur l'ensemble des ω où Z_∞ est strictement positif.*

Démonstration. D'après le Lemme 12.1.3, la suite (Z_n) $(n \geq 0)$ est une martingale *positive*. D'après le Lemme 12.1.1, on a $\mathbb{E}[Z_n] = 1$ pour tout $n \geq 0$, d'où $\sup_{n\geq 0} \mathbb{E}[Z_n] = 1 < +\infty$. La suite (Z_n) $(n \geq 0)$ converge donc presque sûrement vers une variable aléatoire limite, positive et intégrable, en vertu du Théorème 11.1.1. □

Soit maintenant $(p_0, p_1, \ldots)$ la loi de probabilité de X. On se propose d'étudier la nature des états de la chaîne de Markov (Y_n) $(n \geq 0)$ en fonction

de cette loi de probabilité (*cf.* Lemme 12.1.2). Notons $\mathcal{P} = (p_{i,j})$ $((i,j) \in \mathbb{N}^2)$ la matrice de transition de cette chaîne, de sorte que

$$p_{i,j} = \mathrm{P}\{Y_{n+1} = j \mid Y_n = i\} = \mathrm{P}\{X^{(n)}(1) + \cdots + X^{(n)}(i) = j\}.$$

D'abord, si $p_1 = 1$, chaque particule admet, avec une probabilité égale à 1, exactement un descendant, et par conséquent $Y_n = 1$ p.s. Ce cas est inintéressant. On suppose donc, dans la suite, $p_1 < 1$.

Proposition 12.1.5. *L'état* 0 *est absorbant, donc récurrent (extinction).*

Démonstration. Si $Y_n = 0$, alors $Y_{n+1} = 0$, de sorte qu'une fois que le processus atteint 0, il y reste. □

Proposition 12.1.6. *Si* $p_1 < 1$, *alors tout état* $i \geq 1$ *est transient.*

Démonstration. On distingue les cas : $p_0 = 0$, $p_0 = 1$ et $0 < p_0 < 1$.

(a) $p_0 = 0$. Alors chaque particule a au moins un descendant, de sorte que la suite (Y_n) est *croissante.* Si on part de i, à la date n, on retournera au moins une fois dans i, si et seulement si chacun des i membres de la $n^{\text{ième}}$ génération admet *exactement* un descendant. D'où, avec les notations (5.6.5) : $f_{i,i} = \mathrm{P}\{Y_{n+1} = i \mid Y_n = i\} = \mathrm{P}\{X^{(n)}(1) = 1, \ldots, X^{(n)}(i) = 1\} = (p_1)^i < 1$. L'état i est transient.

(b) $p_0 = 1$. Alors $p_{i,0} = 1$, de sorte que $f_{i,i} = 0 < 1$ et i est transient.

(c) $0 < p_0 < 1$. Puisque 0 est absorbant , on a :

$$\begin{aligned} f_{i,i} &\leq \mathrm{P}\{Y_{n+1} \neq 0 \mid Y_n = i\} = 1 - \mathrm{P}\{Y_{n+1} = 0 \mid Y_n = i\} \\ &= 1 - \mathrm{P}\{X^{(n)}(1) = 0, \ldots, X^{(n)}(i) = 0\} = 1 - (p_0)^i < 1 \end{aligned}$$

et i est encore transient. □

Proposition 12.1.7. *Avec une probabilité égale à* 1, *le processus* (Y_n) *s'éteint ou explose, soit* $\mathrm{P}\{Y_n \to 0 \text{ ou } Y_n \to +\infty\} = 1$.

Démonstration. Prenons un nombre entier $N \geq 1$. Puisque les états $1, 2, \ldots, N$ sont tous transients, après un certain temps, le processus (Y_n) ne les visitera plus. Après un certain temps, on aura donc, ou bien $Y_n = 0$, ou bien $Y_n \geq N + 1$. Comme N est arbitraire, on a le résultat. □

12.2 LA LOI DE HARDY-WEINBERG

On considère une population d'individus dont chacun possède une paire spécifiée de gènes, dite *génotype*, chaque gène pouvant être de type A ou a. Initialement, on connaît le pourcentage des individus dont la paire de gènes est de type AA (ou 1), Aa (ou 2), aa (ou 3). En considérant le génotype comme une variable aléatoire, on suppose donc connue la *loi de probabilité initiale du génotype*, qu'on écrit sous la forme : $\mu^{(0)} = (u, 2v, w)$, $u, v, w \geq 0$, $u + 2v + w = 1$.

Quand deux individus se croisent, chacun fournit un de ses deux gènes, avec une probabilité $\frac{1}{2}$ à son descendant. On admet également que, dans un croisement, chaque individu a la même probabilité d'être choisi. La probabilité pour que, lors d'un croisement, un individu pris au hasard fournisse le gène A est $p = u + \frac{1}{2}2v = u + v$ et $q = \frac{1}{2}2v + w = v + w$ pour qu'il fournisse le gène a. On note X_0 le génotype d'un parent initial, X_1 le génotype du descendant de la première génération, ... , X_n le génotype du descendant de la $n^{\text{ième}}$ génération.

On convient que le nombre d'individus est grand, de sorte que si l'on retire un individu quelconque de la population, la proportion des génotypes dans la population restante n'est pas modifiée significativement.

Proposition 12.2.1. *La suite (X_n) $(n \geq 0)$ est une chaîne de Markov, homogène, ayant $E = \{1, 2, 3\}$ comme ensemble des états, $\mu^{(0)}$ comme loi initiale et $\mathcal{P}$ pour matrice de transition, donnée par :*

$$\mathcal{P} = \begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{pmatrix} p & q & 0 \\ p/2 & 1/2 & q/2 \\ 0 & p & q \end{pmatrix} \end{array}.$$

Démonstration. Si $X_0 = 1$, pour avoir $X_1 = 1$, il faut croiser l'individu initial avec un individu de génotype AA (probabilité u) ou un individu de génotype Aa (probabilité $2v\frac{1}{2}$), d'où $p_{1,1} = u + v = p$. Pour avoir $X_1 = 2$, il faut croiser avec un individu de génotype Aa (probabilité $2v\frac{1}{2} = v$) ou un individu de génotype aa (probabilité w), d'où $p_{1,2} = v + w = q$. Ainsi $p_{1,3} = 0$. Si $X_0 = 2$, pour avoir $X_1 = 1$, il faut croiser avec un individu de génotype AA (probabilité $\frac{1}{2}u$) ou un individu de type Aa (probabilité $\frac{1}{2}\frac{1}{2}2v = v/2$. D'où $p_{2,1} = \frac{1}{2}u + \frac{1}{2}v = p/2$. Pour avoir $X_1 = 2$ à partir de $X_0 = 2$, il faut croiser avec un individu de génotype AA (probabilité $\frac{1}{2}u$), ou avec un individu de génotype Aa (probabilté $2\frac{1}{2}\frac{1}{2}2v = v$), ou encore avec un individu de génotype aa (probabilité $\frac{1}{2}w$), d'où $p_{2,2} = \frac{1}{2}u + v + \frac{1}{2}w = \frac{1}{2}$. Les autres coefficients s'obtiennent par symétrie. □

Proposition 12.2.2 (Loi de Hardy-Weinberg). *Pour tout $n \geq 0$, soient X_n le génotype à la $n^{\text{ième}}$ génération et $\mu^{(n)}$ la loi de probabilité de X_n. Alors, quelle que soit la loi de probabilité initiale $\mu^{(0)} = (u, 2v, w)$, pour tout $n \geq 1$, on a $\mu^{(n)} = (p^2, 2pq, q^2)$, avec $p = u + v$, $q = v + w$. En d'autres termes, la loi de probabilité du génotype est stationnaire à partir de la première génération.*

Démonstration. En effet, $\mu^{(1)} = \mu^{(0)}\mathcal{P} = (p^2, 2pq, q^2)$ et lorsque l'on itère encore une fois : $(p^2, 2pq, q^2)\mathcal{P} = (p^2, 2pq, q^2)$. □

12.3 LE BATTAGE DES CARTES

Considérons un jeu de n cartes et convenons que l'ordre initial des n cartes correspond à la permutation identique $(1, 2, \ldots, n)$. Après battage, le paquet se trouve dans un autre ordre, caractérisé par une permutation $\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_n)$ de la suite initiale. Chaque battage fait donc passer d'une permutation σ à une permutation τ. On convient que ces $n!$ différentes permutations sont les $n!$ *états* dans lesquels le paquet de cartes peut se trouver. Une *transition* d'un état σ à un état τ est définie de la façon suivante : on prend la carte située au-dessus du paquet et on l'insère dans le paquet restant, dans l'une des n places possibles. Il y a donc exactement n transitions possibles. On convient que cette insertion dans une place donnée a une probabilité égale à $1/n$. Autrement dit, les seuls états, atteints en une transition à partir de σ, sont les états, notés $f_i(\sigma)$ $(1 \le i \le n)$, donnés par :

$$\begin{aligned}
f_1(\sigma) &= (\sigma_1, \ldots, \sigma_n) = \sigma;\\
f_2(\sigma) &= (\sigma_2, \sigma_1, \sigma_3, \ldots, \sigma_n);\\
\ldots & \quad \ldots\\
f_i(\sigma) &= (\sigma_2, \ldots, \sigma_i, \sigma_1, \sigma_{i+1}, \ldots, \sigma_n);\\
\ldots & \quad \ldots\\
f_n(\sigma) &= (\sigma_2, \ldots, \sigma_n, \sigma_1).
\end{aligned}$$

L'ensemble des états étant l'ensemble E de toutes les $n!$ permutations de $(1, 2, \ldots, n)$, on forme la matrice $\mathcal{P} = (p_{\sigma,\tau})$ $((\sigma, \tau) \in E^2)$, où $p_{\sigma,\tau} = 1/n$ si $\tau = f_i(\sigma)$ pour un certain $i = 1, 2, \ldots, n$ et $p_{\sigma,\tau} = 0$, autrement. On définit ainsi une chaîne de Markov ayant E pour ensemble d'états et $\mathcal{P}$ pour matrice de transition.

Proposition 12.3.1. *La matrice $\mathcal{P}$ est bistochastique.*

Démonstration. Soit σ une permutation ; il y a exactement n transitions $\sigma \mapsto \tau$ de probabilité non-nulle ; ce sont les n *transitions* $\sigma \mapsto f_i(\sigma)$ $(i = 1, 2, \ldots, n)$, de probabilité $1/n$. Chaque *ligne* de $\mathcal{P}$ (de longueur $n!$) contient ainsi exactement n coefficients non-nuls, tous égaux à $1/n$

Maintenant, soit $\tau = (\tau_1, \ldots, \tau_n)$ une permutation. Introduisons, pour tout $i = 2, \ldots, n$, l'application inverse

$$g_i(\tau) := f_i^{-1}(\tau) = (\tau_i, \tau_1, \ldots, \tau_{i-1}, \tau_{i+1}, \ldots, \tau_n).$$

Il y a exactement n transitions $\sigma \mapsto \tau$, de probabilité non-nulle ; ce sont les transitions $g_i(\tau) \mapsto \tau$ $(i = 1, 2, \ldots, n)$, de probabilité $1/n$. Chaque *colonne* de $\mathcal{P}$ a donc exactement n coefficients non nuls, tous égaux à $1/n$. □

Proposition 12.3.2. *La chaîne est irréductible (deux états quelconques communiquent) et acyclique.*

Démonstration. Étant données deux permutations, on sait que l'on peut passer de l'une à l'autre par une suite de *transpositions de deux éléments adjacents.* Or, une telle transposition peut être engendrée par une suite de transitions.

En effet, soit $\sigma = (\sigma_1, \ldots, \sigma_i, \sigma_{i+1}, \ldots, \sigma_n)$ une permutation, que l'on récrit, pour simplifier, $\sigma = A\sigma_i\sigma_{i+1}B$. En posant toujours $g_i = f_i^{-1}$, on a $g_i(\sigma) = \sigma_i A\sigma_{i+1}B$, d'où $(f_{i+1} \circ g_i)(\sigma) = f_{i+1}(g_i(\sigma)) = A\sigma_{i+1}\sigma_i B$. Remarquons enfin que, si l'on applique i fois la transition f_i à une permutation, on retrouve cette permutation ; autrement dit, $(f_i)^i = \mathrm{Id}$, d'où $g_i = (f_i)^{i-1}$ et $f_{i+1} \circ (f_i)^{i-1}(\sigma) = A\sigma_{i+1}\sigma_i B$. Le processus est bien irréductible. Enfin, puisque $p_{\sigma,\sigma} = 1/n$ pour toute permutation σ, le processus est acyclique. □

Proposition 12.3.3. *La chaîne a une loi de probablité stationnaire $u = (u_1, \ldots, u_{n!})$ et une seule, donnée par $u_i = 1/n!$ pour tout $i = 1, \ldots, n!$*

Démonstration. Comme la matrice $\mathcal{P}$ est bistochastique et que le processus est irréductible et acyclique, on peut appliquer les conclusions du Théorème 6.7.1. Pour tout couple (σ, τ), on a $p_{\sigma,\tau}^{(n)} \to 1/n!$ et la loi uniforme sur les $n!$ états est une loi de probabilité stationnaire. En particulier, la fréquence moyenne de séjour dans chaque état est $1/n!$ et le temps moyen de retour dans chaque état est $n!$ □

12.4 LE PROBLÈME DU SCRUTIN

Il s'agit du problème classique du scrutin, que l'on résoud traditionnellement à l'aide du principe de réflexion de Désiré André (*cf.* [FF1], chap. 4, § 5) : *dans un scrutin, on trouve a bulletins pour le candidat A et b bulletins pour le candidat B. On suppose $a > b$. Alors la probabilité pour que, durant tout le dépouillement, A soit toujours en tête est égale à $(a-b)/(a+b)$.*

Dans ce paragraphe, on retrouve, certes, cette évaluation, mais, de façon plus essentielle, on calcule explicitement, à l'aide de la théorie des martingales, des probabilités liées au cheminement aléatoire dans le plan.

Proposition 12.4.1. *Soit (Y_n) $(n \geq 1)$ une suite de variables aléatoires indépendantes, identiquement distribuées, de loi commune $q\varepsilon_{-1} + p\varepsilon_{+1}$. Pour $n \geq 1$, on pose $S_n := Y_1 + \cdots + Y_n$ et l'on désigne par N et c deux entiers supérieurs ou égaux à 2. Alors*

$$\mathrm{P}\{S_1 > 0, \ldots, S_{N-1} > 0 \,|\, S_N = c\} = \frac{c}{N}; \tag{12.4.1}$$

$$\mathrm{P}\{S_1 < c, \ldots, S_{N-1} < c \,|\, S_N = c\} = \frac{c}{N}. \tag{12.4.2}$$

Démonstration. Posons : $X_0 := S_N/N$, $X_1 := S_{N-1}/(N-1)$, … , $X_{N-1} := S_1/1$. On sait (*cf.* Exercice 7.11) que cette suite est une martingale par rapport à (Y_n). L'évènement $A := \{S_1 > 0, \ldots, S_{N-1} > 0\}$ peut encore s'écrire $\{X_1 > 0, \ldots, X_{N-1} > 0\}$. Si A n'est pas réalisé, soit k le premier indice tel que $1 \leq k \leq N-1$ et $X_k = 0$. On a forcément $1 \leq k \leq N-2$, car l'évènement $\{X_{N-1} = 0\}$ ($= \{Y_1 = 0\}$) est impossible, puisque Y_1 ne prend que les valeurs -1 ou $+1$. On pose alors $T := k$ et on définit $T := N-1$ si

l'évènement A est réalisé. La variable aléatoire T ainsi définie est un temps d'arrêt par rapport à la suite (Y_n).

Appliquons à la martingale $X_0, X_1, \ldots, X_{N-1}$ le théorème d'arrêt pour le temps d'arrêt T ; écrivons, par conséquent, $\mathbb{E}[X_0] = \mathbb{E}[X_T]$. Le premier membre de cette identité vaut $\mathbb{E}[S_N]/N$. D'autre part, la variable aléatoire X_T prend la valeur 0 lorsque $T = 1, \ldots, N-2$ et la même valeur que $X_{N-1} = S_1/1 = Y_1$, lorsque $T = N-1$. Comme dans ce dernier cas, A est réalisé, la valeur prise est strictement positive et ne peut donc être que 1. On a donc $\mathbb{E}[X_T] = 1 \cdot \mathrm{P}\{T = N-1\} = \mathrm{P}(A)$ et par suite la relation $\mathrm{P}(A) = \mathbb{E}[S_N]/N$. Récrivons cette relation avec la mesure de probabilité $\mathrm{P}\{\cdot \,|\, S_N = c\}$. On obtient : $\mathrm{P}\{A \,|\, S_N = c\} = c/N$, qui est l'identité (4.1).

Pour obtenir (4.2), observons que, conditionnellement à l'évènement $\{S_N = c\}$, la loi de $c - S_k$ est la même que celle de $S_N - S_k$, qui est la même que celle de S_{N-k}. Il en résulte :

$$\begin{aligned}\mathrm{P}\{A \,|\, S_N = c\} &= \mathrm{P}\{S_1 > 0, \ldots, S_{N-1} > 0 \,|\, S_N = c\}\\ &= \mathrm{P}\{S_1 < c, \ldots, S_{N-1} < c \,|\, S_N = c\}. \quad \square\end{aligned}$$

Retrouvons la formule du scrutin telle qu'elle a été décrite au début de ce paragraphe. Pour ce faire, associons au $n^{\text{ième}}$ votant la variable aléatoire Y_n, qui vaut 1 si le $n^{\text{ième}}$ votant vote pour A et -1 s'il vote pour B. On suppose les Y_n indépendants et identiquement distribués. La quantité $S_n := Y_1 + \cdots + Y_n$ représente la *différence* de voix obtenues par A et B, après le vote du $n^{\text{ième}}$ votant. Avec $N = a + b$ votants et $c = a - b$, la formule (4.1) fournit, conditionnellement à $\{S_N = a - b\}$, la probabilité que cette différence reste strictement positive durant tout le scrutin. On retrouve bien $(a-b)/(a+b)$.

12.5 LE PROBLÈME DU COLLECTIONNEUR ET DE SES FRÈRES

Pierre achète des tablettes de chocolat d'une certaine marque ; dans chaque tablette, il trouve une vignette qu'il colle dans un album édité à cet effet. L'album contient m emplacements, qui correspondent aux m différents types de vignette. On note T le nombre *minimum* de vignettes que Pierre doit acheter pour compléter son album. On suppose qu'à chaque achat, la probabilité d'obtenir une vignette donnée est $1/m$. Pierre a r frères, numérotés de 1 à r, qui eux aussi possèdent chacun un album analogue, mais qui n'achètent pas de tablettes. Si Pierre obtient une vignette dont il a déjà le type, il la donne au frère numéroté 1. Si ce dernier a déjà ce type de vignette dans son album, il la donne au frère numéroté 2 et ainsi de suite. À l'instant T, les frères $1, 2, \ldots, r$ ont respectivement $N^{(1)}$, $N^{(2)}$, … , $N^{(r)}$ emplacements *vides* dans leurs albums. Il s'agit de trouver la valeur moyenne de ces nombres.

Théorème 12.5.1. *On a* $\mathbb{E}[T] = m\,H_m$ *et* $\mathbb{E}[N^{(1)}] = H_m$, *où* H_m *est le* $m^{\text{ième}}$ *nombre harmonique* $1 + \dfrac{1}{2} + \cdots + \dfrac{1}{m}$.

L'évaluation de $\mathbb{E}[T]$ est classique. Elle a fait l'objet de l'Exercice 2.6 et de l'Exercice 6.8. Rappelons que T est la somme de m variables aléatoires géométriques, indépendantes, de paramètres respectifs m/m, $(m-1)/m$, ... , $1/m$. Par conséquent, T est presque sûrement finie. L'évaluation de $\mathbb{E}[N^{(1)}]$ a été imaginée par Pintacuda[1] et résolue par la théorie des martingales. Nous reproduisons ici cette solution, qui nous a été transmise par G. Letta.

Les types de vignettes sont numérotés de 1 à m. On note Y_1, Y_2, ... les types des vignettes que trouve le collectionneur dans ses achats successifs. Par convention, $Y_0 = 0$. On suppose donc que les Y_n $(n \geq 1)$ sont des variables aléatoires indépendantes, chacune uniformément répartie sur $\{1, 2, \dots, m\}$. Pour chaque $n = 0, 1, \dots$, on note $X_n^{(0)}$ le nombre d'emplacements vides dans l'album du collectionneur à l'instant n, de sorte que $X_0^{(0)} = m$, puis $T = \inf\{n \geq 0 : X_n^{(0)} = 0\}$ et enfin $X_T^{(0)} = 0$. Pour $k = 1, 2, \dots$, on note $X_n^{(k)}$ le nombre de vignettes qui sont apparues exactement k fois, jusqu'à la date n incluse. Alors $\sum\limits_{k \geq 1} X_T^{(k)} = m$. De plus, le nombre $N^{(k)}$ d'emplacements vides dans l'album du frère d'indice $k \geq 1$, à cet instant T, est égal à : $N^{(k)} = X_T^{(1)} + \cdots + X_T^{(k)}$.

12.5.1 Le calcul par la méthode des martingales

Supposons donnée une fonction f de $\mathbb{N}^2$ dans $\mathbb{N}$ satisfaisant, pour $x^{(0)} \geq 1$, l'identité

$$x^{(0)}\big(f(x^{(0)} - 1, x^{(1)} + 1) - f(x^{(0)}, x^{(1)})\big) \\ + x^{(1)}\big(f(x^{(0)}, x^{(1)} - 1) - f(x^{(0)}, x^{(1)})\big) = 0 \quad (12.5.1)$$

et formons le processus de terme général $Z_n := f(X_n^{(0)}, X_n^{(1)})$ $(n \geq 0)$. Notons, enfin, $W := Z^{\mid T}$ le *processus arrêté* en T défini par :

$$W_n := Z_{T \wedge n} = Z_n \, I_{\{n<T\}} + Z_T \, I_{\{n \geq T\}}. \quad (12.5.2)$$

Notons que le processus $Z = (Z_n)$ n'est pas une martingale. Il s'agit, en effet, d'un processus borné, positif, qui tend 0 à l'infini, sans être identiquement nul. Or, si c'était une martingale, il serait identiquement nul.

Lemme 12.5.2. *Le processus W est une martingale par rapport à la suite* (Y_n) $(n \geq 0)$.

Démonstration. Désignons par ΔW_n, ... les accroissements $W_{n+1} - W_n$, ... Il s'agit de montrer que $\mathbb{E}[\Delta W_n \mid Y_0, Y_1, \dots, Y_n] = 0$. Montrons d'abord

$$\Delta W_n = \Delta Z_n \, I_{\{X_n^{(0)} > 0\}}. \quad (12.5.3)$$

En effet, $\Delta W_n = W_{n+1} - W_n = Z_{n+1} \, I_{\{n+1<T\}} + Z_T \, I_{\{n+1 \geq T\}} - Z_n \, I_{\{n<T\}} - Z_T \, I_{\{n \geq T\}} = \Delta Z_n \, I_{\{n<T\}} + Z_{n+1}(I_{\{n+1<T\}} - I_{\{n<T\}}) + Z_T(I_{\{n+1 \geq T\}} - I_{\{n \geq T\}}) =$

[1] N. Pintacuda. — Coupon collectors via the martingales, *Boll. Un. Mat. Ital.*, A **17**, (1980), p. 174-177.

$\Delta Z_n\, I_{\{n<T\}} = \Delta Z_n\, I_{\{X_n^{(0)}>0\}}$. L'accroissement des variables W_n est ainsi le même que celui des Z_n, tant que l'album de Pierre n'est pas terminé.

Montrons ensuite que les trois évènements $\{\Delta\, X_n^{(0)} = -1\}$, $\{\Delta\, X_n^{(1)} = -1\}$ et $\{\Delta\, X_n^{(0)} = \Delta\, X_n^{(1)} = 0\}$ forment un système complet d'évènements. En effet, lorsque l'on passe du $n^{\text{ième}}$ achat au $(n+1)^{\text{ième}}$, il n'y a que les trois possibilités suivantes : Pierre obtient une *nouvelle* vignette (l'évènement $\{\Delta X_n^{(0)} = -1\}$ se réalise), ou c'est le cadet ($\{\Delta X_n^{(1)} = -1\}$ se réalise), ou enfin Pierre *et* le cadet ont déjà la vignette qui apparaît au $(n+1)^{\text{ième}}$ achat ($\{\Delta\, X_n^{(0)} = \Delta X_n^{(1)} = 0\}$ se réalise).

Il en résulte

$$\Delta Z_n = I_{\{\Delta X_n^{(0)}=-1\}}(f(X_n^{(0)}-1, X_n^{(1)}+1) - f(X_n^{(0)}, X_n^{(1)})) \\ + I_{\{\Delta X_n^{(1)}=-1\}}(f(X_n^{(0)}, X_n^{(1)}-1) - f(X_n^{(0)}, X_n^{(1)})),$$

On a, d'autre part, $\mathbb{E}[I_{\{\Delta X_n^{(i)}=-1\}}\,|\,Y_0,\ldots,Y_n] = X_n^{(i)}/m$ pour $i = 0, 1$. Prenons l'espérance conditionnelle de ΔZ_n, en tenant compte de la Règle 4 du chap. 1 ; on obtient :

$$\mathbb{E}[\Delta Z_n\,|\,Y_0, Y_1,\ldots,Y_n] = \frac{X_n^{(0)}}{m}(f(X_n^{(0)}-1, X_n^{(1)}+1) - f(X_n^{(0)}, X_n^{(1)})) \\ + \frac{X_n^{(1)}}{m}(f(X_n^{(0)}, X_n^{(1)}-1) - f(X_n^{(0)}, X_n^{(1)})).$$

D'après (12.5.3), on en déduit :

$$\mathbb{E}[\Delta W_n\,|\,Y_0, Y_1,\ldots,Y_n] = I_{\{X_n^{(0)}>0\}}\,\mathbb{E}[\Delta Z_n\,|\,Y_0, Y_1,\ldots,Y_n],$$

une expression qui est nulle d'après l'hypothèse (12.5.1) sur la fonction f. □

Appliquons alors le théorème (d'arrêt) 8.3.1, à la martingale W et au temps d'arrêt T, ce qui est licite, puisque la martingale W est évidemment bornée (elle ne prend qu'un nombre fini de valeurs) et que T est presque sûrement fini : $\mathbb{E}[W_T] = \mathbb{E}[W_0]$. Or, $\mathbb{E}[W_T] = \mathbb{E}[f(X_T^{(0)}, X_T^{(1)})] = \mathbb{E}[f(0, X_T^{(1)})]$ et $\mathbb{E}[W_0] = \mathbb{E}[Z_0] = \mathbb{E}[f(X_0^{(0)}, X_0^{(1)})] = \mathbb{E}[f(m, 0)] = f(m, 0)$. On a ainsi : $\mathbb{E}[f(0, X_T^{(1)})] = f(m, 0)$. Pour obtenir la relation $\mathbb{E}[X_T^{(1)}] = H_m$, il suffit de trouver une fonction f, qui, en plus de (12.5.1), satisfait les propriétés : $f(0, x^{(1)}) = x^{(1)}$ et $f(m, 0) = H_m$. Pintacuda (*op. cit.*) a judicieusement imaginé la fonction : $f(x^{(0)}, x^{(1)}) = H_{x^{(0)}} + \dfrac{x^{(1)}}{1+x^{(0)}}$.

12.5.2 Le calcul par la méthode combinatoire

La méthode de Pintacuda peut aussi être utilisée pour le calcul de $\mathbb{E}[N^{(k)}]$ pour n'importe quel $k \geq 1$, mais a l'inconvénient, même pour $k = 1$, de ne fournir qu'une équation aux différences, dont il faut deviner la solution. Pour

le cas général, nous préférons recourir à une méthode combinatoire, qui a l'avantage de fournir la solution directe et même de donner l'expression des fonctions génératrices sous-jacentes.

La suite des *nombres hyperharmoniques* $K_m^{(k)}$ ($k \geq 0$, $m \geq 1$) est définie par les relations

$$K_1^{(k)} := 0, \text{ pour tout } k \geq 1; \quad K_m^{(0)} := 1, \text{ pour tout } m \geq 1; \quad (12.5.4)$$

et pour $k \geq 1$ et $m \geq 2$ par

$$K_m^{(k)} - \frac{K_m^{(k-1)}}{m} = K_{m-1}^{(k)}. \qquad (12.5.5)$$

Comme tous les nombres $K_m^{(0)}$ valent 1, on voit que $1 + K_m^{(1)}$ n'est autre que le $m^{\text{ième}}$ *nombre harmonique* H_m. La table des premiers nombres hyperharmoniques est reproduite en fin de chapitre.

En utilisant les séries génératrices, la récurrence (12.5.5) s'écrit sous la forme :

$$\Big(1 - \frac{t}{m}\Big) \cdot \sum_{k \geq 0} K_m^{(k)} t^k = \sum_{k \geq 0} K_{m-1}^{(k)} t^k.$$

Puisque $\sum_{k \geq 0} K_1^{(k)} t^k = 1$, on obtient :

$$\sum_{k \geq 0} K_m^{(k)} t^k = \frac{1}{(1 - t/2)(1 - t/3) \cdots (1 - t/m)}. \qquad (12.5.6)$$

En décomposant la fraction rationnelle en éléments simples, on déduit, pour $k \geq 0$ et $m \geq 2$, la formule explicite

$$K_m^{(k)} = m(m-1) \sum_{0 \leq n \leq m-2} \frac{(-m+2)_n}{n!} \frac{1}{(n+2)^{k+1}}, \qquad (12.5.7)$$

où $(a)_n$ désigne la *factorielle montante* :

$$(a)_n := \begin{cases} 1, & \text{si } n = 0; \\ a(a+1) \cdots (a+n-1), & \text{si } n \geq 1. \end{cases}$$

Théorème 12.5.3. *On a* $\mathbb{E}[X^{(1)}] = 1 + K_m^{(1)} = H_m$, *puis* $\mathbb{E}[X^{(k)}] = K_m^{(k)}$ *pour* $k \geq 2$, *soit, par conséquent,* $\mathbb{E}[N^{(k)}] = 1 + K_m^{(1)} + K_m^{(2)} + \cdots + K_m^{(k)}$ *pour tout* $k \geq 1$.

Soit $(t, s_1, s_2, \ldots)$ une suite de variables commutatives. Si f envoie un ensemble fini A dans un ensemble (fini) B d'entiers, on note $\nu_i(f)$ le nombre d'éléments $j \in B$ tels que l'image réciproque $f^{-1}(j)$ soit de cardinal i. Le *poids* $\pi(f)$ de f est défini comme étant le monôme $\pi(f) := \prod_{i \geq 1} s_i^{\nu_i(f)}$.

Développons la série $\Big(\sum_{i\geq 1} s_i \frac{t^i}{i!}\Big)^m$ en une série de puissances de t. On obtient : $\sum_{n\geq m} \frac{t^n}{n!} \sum_{\substack{i_1\geq 1,\ldots,i_m\geq 1\\ i_1+\cdots+i_m=n}} \binom{n}{i_1,\ldots,i_m} s_{i_1}\cdots s_{i_m}$. Comme $i_1 \geq 1, \ldots, i_m \geq 1$, le coefficient multinomial $\binom{n}{i_1,\ldots,i_m}$ n'est autre que le nombre de *surjections* f de $[n] = \{1,\ldots,n\}$ sur $[m] = \{1,\ldots,m\}$ telles que le cardinal de $f^{-1}(1)$ est i_1, … , le cardinal de $f^{-1}(m)$ est i_m. Si on note $\mathrm{Surj}([n],[m])$ l'ensemble des surjections de $[n]$ sur $[m]$, on a donc l'identité :

$$\Big(\sum_{i\geq 1} s_i \frac{t^i}{i!}\Big)^m = \sum_{n\geq m} \frac{t^n}{n!} \sum_{f\in \mathrm{Surj}([n],[m])} \pi(f). \tag{12.5.8}$$

Notons maintenant $A(n,m)$ le sous-ensemble des surjections f appartenant à $\mathrm{Surj}([n],[m])$ telles que si $f(n) = j$, alors j n'apparaît pas dans la suite $(f(1),\ldots,f(n-1))$. Comme la restriction $f\,|_{[n-1]}$ de f à $[n-1]$ est elle-même une surjection de $[n-1]$ sur $[m]\setminus\{f(n)\}$, l'identité (12.5.8) implique

$$ms_1\Big(\sum_{i\geq 1} s_i \frac{t^i}{i!}\Big)^{m-1} = \sum_{n-1\geq m-1} \frac{t^{n-1}}{(n-1)!} \sum_{f\in A(n,m)} \pi(f). \tag{12.5.9}$$

Soient alors $r \geq 1$, puis $\mathbf{n} := (n_1, n_2, \ldots, n_r)$ une suite d'entiers positifs et enfin $A(n,m;\mathbf{n})$ le sous-ensemble des surjections f appartenant à $A(n,m)$ telles que $\nu_i(f) = n_i$ pour $i = 1, 2, \ldots, r$.

En posant $s_i := 1$ pour tous les entiers $i \geq r+1$ et en notant $\#A(n,m;\mathbf{n})$ le cardinal de $A(n,m;\mathbf{n})$, l'identité (12.5.9) se récrit :

$$\begin{aligned} &ms_1\Big(e^t - 1 + t(s_1-1) + \frac{t^2}{2!}(s_2-1) + \cdots + \frac{t^r}{r!}(s_r-1)\Big)^{m-1} \\ &\qquad = \sum_{n\geq m} \frac{t^{n-1}}{(n-1)!} \sum_{\mathbf{n}} \#A(n,m;\mathbf{n}) \prod_{1\leq i\leq r} s_i^{n_i}. \end{aligned} \tag{12.5.10}$$

Maintenant, l'évènement $\{T = n, X_T^{(1)} = n_1, \ldots, X_T^{(r)} = n_r\}$ se réalise si et seulement si l'application $f : i \mapsto Y_i$ $(1 \leq i \leq n)$ appartient à l'ensemble $A(n,m;\mathbf{n})$. De là,

$$\mathrm{P}\{T = n, X_T^{(1)} = n_1, \ldots, X_T^{(r)} = n_r\} = \frac{\#A(n,m;\mathbf{n})}{m^n}.$$

La *fonction génératrice* du vecteur aléatoire $(T, X_T^{(1)}, \ldots, X_T^{(r)})$ est donc donnée par

$$\begin{aligned} &\sum_{n,\mathbf{n}} \mathrm{P}\{T = n, X_T^{(1)} = n_1, \ldots, X_T^{(r)} = n_r\} t^n s_1^{n_1}\cdots s_r^{n_r} \\ &\qquad = \sum_{n\geq m} \frac{t^n}{m^n} \sum_{\mathbf{n}} \#A(n,m;\mathbf{n}) \prod_{1\leq i\leq r} s_i^{n_i} \end{aligned} \tag{12.5.11}$$

Or le membre de droite de (12.5.10) est une série normalisée par des factorielles $n!$ et non par des puissances m^n. Il nous faut donc introduire l'opérateur Ω_m, qui envoie une série de la forme $a = \sum_{n\geq 0} a_n(s_1,\ldots,s_r)t^n/n!$ sur la série $\Omega_m\, a := \sum_{n\geq 0} a_n(s_1,\ldots,s_r)t^n/m^n$. On vérifie, très facilement, que la série transformée peut encore s'écrire :

$$\Omega_m\, a = m\int_0^{+\infty} e^{-mx}\, a_n(xs_1,\ldots,xs_r)\,dx. \qquad (12.5.12)$$

Appliquons l'opérateur Ω_m à l'identité (12.5.10) et multiplions ensuite les deux membres par t/m ; on obtient

$$t\,\Omega_m\Big(s_1\Big(e^t - 1 + t(s_1-1) + \frac{t^2}{2!}(s_2-1) + \cdots + \frac{t^r}{r!}(s_r-1)\Big)^{m-1}$$
$$= \sum_{n\geq m}\frac{t^n}{m^n}\sum_{\mathbf{n}} \#A(n,m;\mathbf{n}) \prod_{1\leq i\leq r} s_i^{n_i},$$

soit, en tenant compte de (12.5.11) et (12.5.12),

$$(12.5.13)\quad \sum_{n,\mathbf{n}} \mathrm{P}\{T=n, X_T^{(1)}=n_1,\ldots,X_T^{(r)}=n_r\}t^n s_1^{n_1}\cdots s_r^{n_r}$$
$$= ms_1 t\int_0^{+\infty} e^{-mx}\Big(e^{xt}-1+\frac{xt}{1!}(s_1-1)+\cdots+\frac{(xt)^r}{r!}(s_r-1)\Big)^{m-1}dx.$$

On obtient ainsi, sous la forme d'une intégrale, la fonction génératrice du vecteur aléatoire $(T, X_T^{(1)},\ldots,X_T^{(r)})$ Cependant, en développant la fonction à intégrer au moyen du théorème multinomial, on peut calculer explicitement cette intégrale ; on trouve :

$$s_1 t\sum \binom{m-1}{a,b,c_1,\ldots,c_r}(-1)^b\Big(\prod_{k=1}^r\Big(\frac{s_k-1}{k!}\Big)^{c_k}\Big)\frac{(t/m)^{\Sigma kc_k}}{(1-at/m)^{1+\Sigma kc_k}}\big(\sum_k kc_k\big)!$$

De cette expression, on retrouve la fonction génératrice usuelle de T et donc immédiatement son espérance $\mathbb{E}[T] = m\,H_m$. Le plus important est qu'une fois obtenues les fonctions génératrices des variables $X_T^{(k)}$ ($k\geq 1$), on puisse, grâce à la formule explicite (12.5.7), *sommer* leurs dérivées en 1, et ainsi déterminer leurs espérances. En effet, on déduit :

$$\sum_{n\geq 1}\mathrm{P}\{X_T^{(1)}=n\}s^n = s\sum_{a+b+c=m-1}\binom{m-1}{a,b,c}(-1)^b\frac{m(s-1)^c}{(m-a)^{1+c}}c!\,;$$

et pour $k\geq 2$

$$\sum_{n\geq 1}\mathrm{P}\{X_T^{(k)}=n\}s^n = \sum_{a+b+c=m-1}\binom{m-1}{a,b,c}(-1)^b\frac{(s-1)^c}{k!^c}\frac{m}{(m-a)^{1+kc}}(kc)!$$

D'où, par dérivation,

$$\mathbb{E}[X_T^{(1)}] = -\sum_{b=0}^{m-1} \frac{(-m)_{b+1}}{(b+1)!} + m(m-1)\sum_{b=0}^{m-2} \frac{(-m+2)_b}{b!}\frac{1}{(b+2)^2}$$
$$= 1 + K_m^{(1)} = H_m,$$

le résultat déjà prouvé dans le précédent paragraphe. De plus, pour $k \geq 2$,

$$\mathbb{E}[X_T^{(k)}] = m(m-1)\sum_{b=0}^{m-2} \frac{(-m+2)_b}{b!}\frac{1}{(b+2)^{1+k}} = K_m^{(k)}.$$

Table des nombres hyperharmoniques :

m	1	2	3	4	5	10	15	20	25	30	35	40	45	50
$K_m^{(0)}$	1	1	1	1	1	1	1	1	1	1	1	1	1	1
$K_m^{(1)}$	0	0,50	0,83	1,08	1,28	1,93	2,31	2,60	2,82	3,00	3,15	3,28	3,39	3,50
$K_m^{(2)}$	0	0,25	0,53	0,80	1,06	2,14	2,98	3,67	4,27	4,79	5,26	5,68	6,07	6,43
$K_m^{(3)}$	0	0,13	0,30	0,50	0,71	1,79	2,82	3,76	4,64	5,46	6,23	6, 96	7,65	8,30
$K_m^{(4)}$	0	0,06	0,16	0,29	0,43	1,27	2,20	3,14	4,08	4,99	5,89	6,77	7,63	8,47
$K_m^{(5)}$	0	0,03	0,09	0,16	0,24	0,81	1,51	2,28	3,08	3,91	4,75	5,59	6,44	7,29
$K_m^{(6)}$	0	0,02	0,04	0,08	0,13	0,48	0,95	1,49	2,09	2,73	3,40	4,09	4,80	5,52

Par exemple, avec $m = 50$, l'album du collectionneur se termine, en moyenne, après 225 achats. Il y a alors, toujours en moyenne, seulement 4 ou 5 places vides dans l'album du premier frère, 11 places vides dans celui du second frère, 19 dans l'album du troisième frère, etc.

Chapitre 13

Le mouvement brownien

Ce dernier chapitre est consacré à l'étude du mouvement brownien. Il n'était pas question de faire une théorie exhaustive. On n'y trouvera que les propriétés essentielles et un exposé sur les temps d'atteinte. On s'est efforcé de faire un parallèle avec les processus à temps discret, lorsque ceci était possible.

13.1 LE MOUVEMENT BROWNIEN COMME LIMITE D'UNE PROMENADE ALÉATOIRE

Considérons une particule qui se déplace sur $\mathbb{Z}$, en pratiquant une *promenade aléatoire symétrique* : dans chaque unité de temps, elle se déplace d'une unité de longueur vers la droite ou vers la gauche avec probabilité $\frac{1}{2}$. On est donc en présence de la chaîne de Markov, homogène sur $\mathbb{Z}$, définie par la matrice de transition $\mathcal{P} = (p_{i,j})$ $(i, j \in \mathbb{Z})$, où $p_{i,i+1} = \frac{1}{2}$, $p_{i,i-1} = \frac{1}{2}$ pour tout $i \in \mathbb{Z}$ et où tous les autres coefficients sont nuls (*cf.* § 5.2.5).

Supposons maintenant que la promenade soit accélérée en prenant des sauts de plus en plus petits, dans des intervalles de temps de plus en plus petits. Y a-t-il un processus limite, obtenu en faisant tendre vers 0 la longueur des sauts et l'intervalle de temps ?

De façon précise, supposons que, dans chaque intervalle de temps, de longueur Δt, la particule se déplace d'une longueur Δx vers la droite ou vers la gauche avec probabilité $\frac{1}{2}$ et notons $X(t)$ sa position à la date $t > 0$ en faisant l'hypothèse $X(0) = 0$. On a alors

$$X(t) = \Delta x \,(X_1 + \cdots + X_{[t/\Delta t]}), \tag{13.1.1}$$

où

$$X_i = \begin{cases} +1, & \text{si le } i^{\text{ième}} \text{ saut de longueur } \Delta x \text{ s'effectue vers la droite ;} \\ -1, & \text{si le } i^{\text{ième}} \text{ saut de longueur } \Delta x \text{ s'effectue vers la gauche.} \end{cases}$$

En supposant les variables aléatoires X_i indépendantes, de même loi, cette loi commune étant la loi de Bernoulli $\frac{1}{2}(\varepsilon_{-1} + \varepsilon_{+1})$, on a $\mathbb{E}[X_i] = 0$ et $\operatorname{Var} X_i = 1$ pour tout i. De (13.1.1), il résulte alors :

$$\mathbb{E}[X(t)] = 0, \qquad \operatorname{Var} X(t) = (\Delta x)^2 \, [t/\Delta t]. \tag{13.1.2}$$

Il s'agit maintenant de passer à la limite, de sorte que le processus limite ne soit pas trivial. Si, par exemple, on prend $\Delta x = \Delta t$ et on fait tendre Δt vers 0, on obtient $\mathbb{E}[X(t)] = 0$ et $\operatorname{Var} X(t) \to 0$. Le processus $X(t)$ serait alors presque sûrement nul.

Prenons $\Delta x = \sqrt{\Delta t}$ et $\Delta t \to 0$. Alors $\mathbb{E}[X(t)] = 0$ et $\operatorname{Var} X(t) \to t$. On obtient un processus limite $(X(t))$ $(t \geq 0)$ admettant les propriétés suivantes :

Propriété 13.1.1. *Pour $t > 0$ fixé, la variable aléatoire $X(t)$ suit la loi normale $N(0, \sqrt{t})$.*

Démonstration. En effet, d'après (13.1.1), la variable $X(t)$ réduite s'écrit :

$$\frac{\Delta x (X_1 + \cdots + X_{[t/\Delta t]})}{\sqrt{t}} = \frac{X_1 + \cdots + X_{[t/\Delta t]}}{\sqrt{t/\Delta t}} = \frac{X_1 + \cdots + X_{[t/\Delta t]}}{\sigma(X_1 + \cdots + X_{[t/\Delta t]})}.$$

D'après le théorème « central limit » (*cf.* [FF1]; chap. 1, § 11), cette quantité tend en loi, vers une variable aléatoire de loi $N(0,1)$. □

Propriété 13.1.2. *Le processus $\{X(t) : t \geq 0\}$ est à accroissements indépendants, ce qui veut dire que pout toute suite $0 < t_1 < \cdots < t_n$ $(n \geq 2)$, les variables aléatoires $X(t_1)$, $X(t_2) - X(t_1)$, … , $X(t_n) - X(t_{n-1})$ sont indépendantes.*

Démonstration. Cet énoncé résulte du fait que, dans la promenade aléatoire, les sauts ayant lieu dans des intervalles de temps disjoints sont indépendants. □

Propriété 13.1.3. *Le processus $\{X(t) : t \geq 0\}$ est à accroissements stationnaires, ce qui signifie que la distribution de la variable $X(t+s) - X(t)$ ne dépend pas de t.*

Démonstration. Cette propriété résulte du fait que, dans la promenade aléatoire, le changement de position dans un intervalle de temps ne dépend que de la longueur de cet intervalle de temps. □

Cette introduction heuristique permet de donner la définition du mouvement brownien, comme développé dans le prochain paragraphe.

13.2 DÉFINITIONS ET PREMIÈRES PROPRIÉTÉS

Définition 13.2.1. Un processus stochastique $\{X(t) : t \geq 0\}$ est appelé *mouvement brownien (standard)* si les trois conditions suivantes sont satisfaites :

(1) $X(0) = 0$;

(2) $\{X(t) : t \geq 0\}$ est *à accroissements indépendants* et *stationnaires*;

(3) pour tout $t > 0$, la variable aléatoire $X(t)$ suit la loi normale $N(0, \sqrt{t})$ (de moyenne nulle et d'*écart-type* $\sqrt{t}$).

Remarque 1. Ce processus est également appelé *processus de Wiener* (1918) ; il a été étudié antérieurement par le botaniste Robert Brown (1827), puis par Einstein (1905).

Remarque 2. Si pour tout $t > 0$, la variable $X(t)$ suit la loi normale $N(0, c\sqrt{t})$ $(c > 0)$, on parle encore de mouvement brownien; si $c = 1$, on dit que c'est le mouvement brownien *standard*. Tout mouvement brownien $\{X(t) : t \geq 0\}$ peut être ramené au mouvement brownien standard en considérant $X(t)/c$; nous supposerons donc toujours $c = 1$.

Propriété 13.2.1. *Avec une probabilité égale à un, $X(t)$ est une fonction continue de t ; autrement dit, presque toute trajectoire du mouvement brownien est continue.*

Nous ne démontrons pas cette propriété, mais nous observons ce qui suit.

(a) *La fonction aléatoire $X(t)$ est localement continue en probabilité.*

En effet, si $h > 0$, on a, pour tout $\varepsilon > 0$

$$P\{|X(t+h) - X(t)| > \varepsilon\} \leq \frac{\mathrm{Var}(X(t+h) - X(t))}{\varepsilon^2} = \frac{\mathrm{Var}\, X(h)}{\varepsilon^2} = \frac{h}{\varepsilon^2},$$

une expression qui tend vers 0 lorsque h tend vers 0. □

(b) *La fonction aléatoire $X(t)$ n'est dérivable en aucun point $t \in \mathbb{R}$.*

En effet, pour $h > 0$, la variable aléatoire $Y = \dfrac{X(t+h) - X(t)}{h}$ est normale, de moyenne nulle et de variance $1/h$. Par conséquent, pour tout $N > 0$,

$$\begin{aligned} P\{|Y| > N\} &= P\{|X(t+h) - X(t)| > Nh\} \\ &= P\{|X(h)| > Nh\} = P\left\{\left|\frac{X(h)}{\sqrt{h}}\right| > N\sqrt{h}\right\}, \end{aligned}$$

d'où, puisque $X(h)/\sqrt{h}$ est de loi $N(0, 1)$,

$$P\{|Y| > N\} = \frac{1}{\sqrt{2\pi}} \int_{|y| > N\sqrt{h}} e^{-y^2/2}\, dy,$$

une expression qui tend vers 1 lorsque h tend vers 0; d'où le résultat, N étant arbitraire. □

Propriété 13.2.2 (Loi temporelle du mouvement brownien). *Soit* $0 < t_1 < t_2 < \cdots < t_n < +\infty$. *La loi du vecteur* $(X(t_1), X(t_2), \ldots, X(t_n))$ *est une loi normale à n dimensions, dont la densité conjointe* $f(x_1, x_2, \ldots, x_n)$ *est donnée par*

$$\begin{aligned} f(x_1, x_2, \ldots, x_n) \\ = \frac{1}{(2\pi)^{n/2}\sqrt{t_1(t_2-t_1)\cdots(t_n-t_{n-1})}} \\ \times \exp\Big(-\frac{1}{2}\Big(\frac{x_1^2}{t_1} + \frac{(x_2-x_1)^2}{t_2-t_1} + \cdots + \frac{(x_n-x_{n-1})^2}{t_n-t_{n-1}}\Big)\Big). \quad (13.2.1) \end{aligned}$$

Démonstration. La densité du vecteur $(X(t_1), X(t_2), \ldots, X(t_n))$ au point $(x_1, x_2, \ldots, x_n)$ est la même que celle de $(X(t_1), X(t_2) - X(t_1), \ldots, X(t_n) - X(t_{n-1}))$ au point $(x_1, x_2 - x_1, \ldots, x_n - x_{n-1})$. Or, ce dernier vecteur a toutes ses composantes *indépendantes*, puisque le processus est à accroissements indépendants et toutes *normales*, de moyenne nulle. Enfin, puisque le processus est à accroissements stationnaires, les variances de ces composantes sont t_1, $t_2 - t_1$, … , $t_n - t_{n-1}$, respectivement. □

Propriété 13.2.3. *Le vecteur* $(X(t_1), X(t_2), \ldots, X(t_n))$ *est centré et sa matrice des variances-covariances est donnée par :*

$$\Gamma = \begin{pmatrix} t_1 & t_1 & t_1 & \ldots & t_1 \\ t_1 & t_2 & t_2 & \ldots & t_2 \\ t_1 & t_2 & t_3 & \ldots & t_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ t_1 & t_2 & t_3 & \ldots & t_n \end{pmatrix}.$$

Démonstration. Il est clair que le vecteur est *centré.* On peut retrouver l'expression de la matrice des variances-covariances Γ à partir de la la densité donnée dans la Propriété 13.2.2, mais il est plus intéressant de procéder comme suit.

Si V est un vecteur-colonne, notons ${}^tV$ le *vecteur-ligne* transposé de V. Soient alors ${}^tM'$ le vecteur-ligne

$$(Y_1, Y_2, \ldots, Y_n) := (X(t_1), X(t_2) - X(t_1), \ldots, X(t_n) - X(t_{n-1}))$$

et ${}^tM$ le vecteur-ligne $(X(t_1), X(t_2), \ldots, X(t_n))$, de sorte que le vecteur-colonne M est égal au produit matriciel $A\,M'$, où A est la matrice triangulaire inférieure qui a des 1 sur et au-dessous de la diagonale et des zéros au-dessus. Comme le vecteur M' suit une loi normale à n dimensions $\mathcal{N}(0, \Gamma')$, où la matrice des variances-covariances Γ' est la matrice-diagonale $\operatorname{diag}(t_1, t_2 - t_1, \ldots, t_n - t_{n-1})$, le vecteur M suit la loi normale $\mathcal{N}(0, \Gamma)$, avec $\Gamma = A\,\Gamma'\,{}^tA$ (*cf.*, par exemple, [FF1], Théorème 5.6, p. 153).

On peut naturellement calculer Γ à l'aide de cette dernière formule, mais il est plus instructif d'opérer directement. Pour $1 \leq i \leq j \leq n$, on a : $\operatorname{Cov}(X(t_i), X(t_j)) = \mathbb{E}[X(t_i)X(t_j)] = \mathbb{E}[X(t_i)(X(t_i) + Y_{i+1} + \cdots + Y_j)]$. Comme

$X(t_i)$ est indépendant de $(Y_{i+1} + \cdots + Y_j)$ et que $\mathbb{E}[X(t_i)] = 0$, on en déduit $\mathrm{Cov}(X(t_i), X(t_j)) = \mathbb{E}[X(t_i)^2] = t_i$ et donc

$$\mathrm{Cov}(X(t_i), X(t_j)) = \mathbb{E}[X(t_i)X(t_j)] = t_i \quad (1 \leq i \leq j \leq n); \tag{13.2.2}$$

soit aussi

$$\mathrm{Cov}(X(s), X(t)) = \min(s, t). \quad \square \tag{13.2.3}$$

13.3 TEMPS D'ATTEINTE, VARIABLES ALÉATOIRES MAXIMALES

La propriété que les trajectoires du mouvement brownien sont presque sûrement *continues*, jointe au *principe de réflexion* (*cf.* [FF1], chap. 4, § 5) permet de calculer très simplement des probabilités intéressantes associées au mouvement brownien.

13.3.1 Temps d'atteinte

Pour $a > 0$, désignons par T_a le premier instant où le mouvement brownien atteint a (*cf.* Figure 13.1), soit

$$T_a := \inf\{t \geq 0 : X(t) = a\}. \tag{13.3.1}$$

Admettons que ce temps d'atteinte est presque sûrement fini, de sorte que l'on peut considérer des variables de la forme $X(T_a + t)$ avec $t \geq 0$.

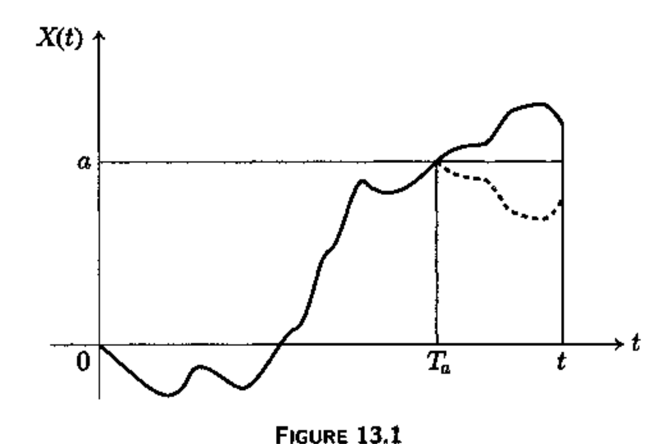


FIGURE 13.1

Définissons le processus $\{Y(t) : t \geq 0\}$ comme étant égal à $\{X(t) : t \geq 0\}$ jusqu'à la date T_a et égal, après cet instant, au symétrique du processus $\{X(t) : t \geq 0\}$ par rapport à la droite d'ordonnée a :

$$Y(t) := \begin{cases} X(t), & \text{si } t < T_a\,; \\ 2\,X(T_a) - X(t) = 2a - X(t), & \text{si } t \geq T_a. \end{cases}$$

Proposition 13.3.1 (Principe de réflexion). *Les processus* $\{X(t) : t \geq 0\}$ *et* $\{Y(t) : t \geq 0\}$ *ont même loi.*

Démonstration. Pour $t < T_a$, on a $X(t) = Y(t)$, par définition. Pour des valeurs supérieures ou égales à T_a, c'est-à-dire pour $T_a + t$ avec $t \geq 0$, écrivons $Y(T_a+t) = 2a - X(T_a+t) = a + (a - X(T_a+t))$. On fait apparaître le processus $\{a - X(T_a+t) : t \geq 0\}$, qui a même loi que le processus $\{X(T_a+t) - a : t \geq 0\}$. Or, ce dernier est le processus qui *redémarre* en 0 à l'instant T_a. Il est intuitif que c'est encore un mouvement brownien et qu'il est indépendant du passé jusqu'à la date T_a, tout comme son symétrique $\{a - X(T_a + t) : t \geq 0\}$.

Pour démontrer ces deux propriétés rigoureusement, il faudrait introduire une notion de propriété de Markov forte pour le mouvement brownien, analogue à celle qui avait été formulée pour les chaînes de Markov au paragraphe 5.5. En admettant ces deux propriétés, on voit donc que $Y(T_a + t) = a + (a - X(T_a + t))$ est indépendant du passé jusqu'à la date T_a et a même loi que $a + (X(T_a + t) - a) = X(T_a + t)$. □

La fonction de répartition de T_a se calcule en considérant $\mathrm{P}\{X(t) \geq x\}$ pour $x \geq a$ et en conditionnant par rapport à l'évènement $\{T_a \leq t\}$ et son contraire. Pour $x \geq a$, on a :

$$\mathrm{P}\{X(t) \geq x\} = \mathrm{P}\{X(t) \geq x \,|\, T_a \leq t\}\,\mathrm{P}\{T_a \leq t\} + \mathrm{P}\{X(t) \geq x \,|\, T_a > t\}\,\mathrm{P}\{T_a > t\}.$$

Si $\{T_a \leq t\}$ est réalisé, alors le processus atteint a, pour la première fois, en un point de l'intervalle $[0, t]$ (*cf.* Fig. 13.1). Avec les notations de la Proposition 13.3.1, on a donc

$$\begin{aligned}\mathrm{P}\{X(t) \geq x \,|\, T_a \leq t\} &= \mathrm{P}\{Y(t) \leq 2a - x \,|\, T_a \leq t\} \\ &= \mathrm{P}\{X(t) \leq 2a - x \,|\, T_a \leq t\}\end{aligned}$$

et par conséquent $\mathrm{P}\{X(t) \geq a \,|\, T_a \leq t\} = \frac{1}{2}$.

Si $\{T_a > t\}$ est réalisé, alors il est clair que $\mathrm{P}\{X(t) \geq a \,|\, T_a > t\} = 0$, car, par continuité, le processus ne peut prendre une valeur supérieure ou égale à x sans avoir, au préalable, atteint a. On a donc finalement : $\mathrm{P}\{X(t) \geq a\} = \frac{1}{2}\,\mathrm{P}\{T_a \leq t\}$; d'où

$$\begin{aligned}\mathrm{P}\{T_a \leq t\} &= 2\,\mathrm{P}\{X(t) \geq a\} \\ &= \frac{2}{\sqrt{2\pi t}} \int_a^{+\infty} e^{-x^2/(2t)}\,dx = \frac{2}{\sqrt{2\pi}} \int_{a/\sqrt{t}}^{+\infty} e^{-y^2/2}\,dy. \quad (13.3.2)\end{aligned}$$

Remarque 1. Comme, pour $a < 0$, la loi de T_a est la même que celle de T_{-a}, on en tire, pour a quelconque,

$$\mathrm{P}\{T_a \leq t\} = \frac{2}{\sqrt{2\pi}} \int_{|a|/\sqrt{t}}^{+\infty} e^{-y^2/2}\,dy.$$

Remarque 2. La *densité* de T_a, pour $a > 0$, s'obtient par dérivation :

$$f_{T_a}(t) = \frac{a}{\sqrt{2\pi}} t^{-3/2} e^{-a^2/(2t)} \quad (t > 0). \tag{13.3.3}$$

13.3.2 Variables aléatoires maximales

Pour $a > 0$, on a, par continuité,

$$\mathrm{P}\{\sup_{0 \le s \le t} X(s) \ge a\} = \mathrm{P}\{T_a \le t\},$$

c'est-à-dire, d'après (13/3.2),

$$\mathrm{P}\{\sup_{0 \le s \le t} X(s) \ge a\} = \frac{2}{\sqrt{2\pi}} \int_{a/\sqrt{t}}^{+\infty} e^{-y^2/2}\, dy.$$

13.3.3 Récurrence

Dans la relation précédente, en gardant $a > 0$, faisons tendre t vers $+\infty$. On obtient :

$$\mathrm{P}\{\sup_{s \ge 0} X(s) \ge a\} = \mathrm{P}\{T_a < +\infty\} = 1.$$

La dernière égalité montre que le processus $\{X(t) : t \ge 0\}$ est *récurrent.*

13.4 LE PONT BROWNIEN

Définition 13.4.1. Soit $\{X(t) : t \ge 0\}$ un mouvement brownien standard. Le processus $\{Y(t) : 0 \le t \le 1\}$, où $Y(t) := X(t) - t\,X(1)$ est appelé *pont brownien.* La fonction $t \mapsto \mathbb{E}[Y(t)]$ $(0 \le t \le 1)$ et la fonction $(s,t) \mapsto \mathrm{Cov}(Y(s), Y(t))$ $(0 \le s \le t)$ sont appelées *fonction d'espérance* et *fonction des variances-covariances.*

Presque toute trajectoire de ce pont est une courbe *continue* sur $[0,1]$ passant par les points $(0,0)$ et $(1,0)$.

Propriété 13.4.1. *Les fonctions d'espérance et des variances-covariances du pont brownien sont données par :*

(a) $\mathbb{E}[Y(t)] = 0$ $(0 \le t \le 1)$;

(b) $\mathrm{Cov}((Y(s), Y(t)) = s(1-t)$ $(0 \le s \le t \le 1)$.

Démonstration. La première identité est triviale; pour prouver la seconde, on fait usage de l'identité (13.2.3) : $\mathrm{Cov}(Y(s), Y(t)) = \mathrm{Cov}(X(s) - sX(1), X(t) - tX(1)) = \mathbb{E}[X(s)X(t)] - s\,\mathbb{E}[X(t)X(1)] - t\,\mathbb{E}[X(s)X(1)] + st\,\mathbb{E}[X^2(1)] = s - st - ts + st = s - st = s(1-t)$. □

La loi du pont brownien, qui vient d'être défini et que l'on peut appeler *pont brownien, par excellence*, admet une interprétation au moyen du conditionnement $\{X(1) = 0\}$, comme décrit dans le théorème suivant. On y utilise les notations suivantes :

$\{X(t) : t \geq 0\}$ est un mouvement brownien standard ;
$0 < t_1 < t_2 < \cdots < t_n < 1$ $(n \geq 1)$ est une suite de nombres ;
le vecteur-ligne $(X(t_1), X(t_2), \ldots, X(t_n))$ est noté ${}^tX$; on pose, de plus,
${}^tY := (Y_1, \ldots, Y_n) := (X(t_1) - t_1X(1), \ldots, X(t_n) - t_nX(1))$;
${}^tZ := (Z_1, \ldots, Z_n) := (t_1X(1), \ldots, t_nX(1))$; de sorte que l'on a l'identité : $X = Y + Z$.

Comme décrit dans le paragraphe 1.13, la loi du vecteur X conditionnelle à $\{X(1) = 0\}$ est définie comme l'espérance conditionnelle $\mathbb{E}[I_{\{X \in B\}} \mid X(1) = 0]$. On peut aussi calculer la probabilité conditionnelle $\mathrm{P}\{X \in B \mid |X(1)| \leq 1/k\}$ et faire tendre ensuite k vers l'infini (*cf.* Exercice 13.9).

Théorème 13.4.2. *La loi du vecteur Y est égale à la loi, conditionnelle à $\{X(1) = 0\}$, du vecteur X.*

Démonstration. Les vecteurs Y et Z suivent évidemment des lois normales à n dimensions. Ils sont, en outre, *indépendants* ; pour le vérifier, il suffit de montrer que $\mathbb{E}[Y\,{}^tZ] = 0$ ou que, pour $1 \leq i, j \leq n$, on a : $\mathbb{E}[(X(t_i) - t_iX(1))\, t_jX(1)] = 0$; ce qui est vrai, puisque cette quantité vaut $t_j\, \mathbb{E}[X(t_i)X(1)] - t_it_j\, \mathbb{E}[X^2(1)] = t_jt_i - t_it_j = 0$.

Par ailleurs, l'évènement $\{X(1) = 0\}$ est réalisé si et seulement si $\{Z = 0\}$ est réalisé. En appliquant la formule (1.13.8), pour toute suite de boréliens $B_1, \ldots, B_n$, on obtient :

$$\begin{aligned}\mathrm{P}\{\bigcap_{i=1}^{n} \{X(t_i) \in B_i\} \mid X(1) = 0\} &= \mathrm{P}\{\bigcap_{i=1}^{n} \{(Y_i + Z_i) \in B_i\} \mid Z = 0\}\} \\ &= \mathrm{P}\{\bigcap_{i=1}^{n} \{Y_i \in B_i\}\}. \quad \square\end{aligned}$$

La proposition suivante est donc aussi démontrée.

Proposition 13.4.3. *Soient $\{X(t) : t \geq 0\}$ un mouvement brownien standard et $0 < t_1 < \cdots < t_n < 1$ une suite de nombres. Alors la loi de $(X(t_1), \ldots, X(t_n))$, conditionnelle à $\{X(1) = 0\}$ suit la loi normale $\mathcal{N}(0, \Gamma')$, où*

$$\Gamma' = \begin{pmatrix} t_1(1-t_1) & t_1(1-t_2) & t_1(1-t_3) & \ldots & t_1(1-t_n) \\ t_1(1-t_2) & t_2(1-t_2) & t_2(1-t_3) & \ldots & t_2(1-t_n) \\ t_1(1-t_3) & t_2(1-t_3) & t_3(1-t_3) & \ldots & t_3(1-t_n) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ t_1(1-t_n) & t_2(1-t_n) & t_3(1-t_n) & \ldots & t_n(1-t_n) \end{pmatrix}.$$

En particulier, pour $0 < s < 1$, la loi de $X(s)$ conditionnelle à $\{X(1) = 0\}$ est la loi normale $N(0, \sqrt{s(1-s)})$.

13.5 QUELQUES MODIFICATIONS DU MOUVEMENT BROWNIEN

Les modifications introduites dans ce paragraphe sont toutes des processus de Markov dont les trajectoires sont presque sûrement continues.

13.5.1 Le mouvement brownien réfléchi

Définition 13.5.1. Soit $\{X(t) : t \geq 0\}$ un mouvement brownien. Le processus $\{Y(t) : t \geq 0\}$, où $Y(t) := |X(t)|$ $(t \geq 0)$, est appelé *mouvement brownien réfléchi* (à l'origine).

Proposition 13.5.1. *On a :* $\mathbb{E}[Y(t)] = \sqrt{2t/\pi}$ *et* $\operatorname{Var} Y(t) = (1 - 2/\pi)t$.

Démonstration. D'abord,

$$\mathbb{E}[Y(t)] = \int_{\mathbb{R}} |x| \, \frac{1}{\sqrt{2\pi t}} \, e^{-x^2/(2t)} \, dx = 2 \int_0^{+\infty} x \frac{1}{\sqrt{2\pi t}} \, e^{-x^2/(2t)} \, dx,$$

d'où le résultat en faisant le changement de variables $x^2/(2t) := u$. Ensuite,

$$\begin{aligned}\operatorname{Var} Y(t) &= \mathbb{E}[Y^2(t)] - (\mathbb{E}[Y(t)])^2 \\ &= \mathbb{E}[X^2(t)] - (\mathbb{E}[Y(t)])^2 = t - \frac{2t}{\pi} = t\Big(1 - \frac{2}{\pi}\Big). \quad \square\end{aligned}$$

13.5.2 Le mouvement brownien absorbé

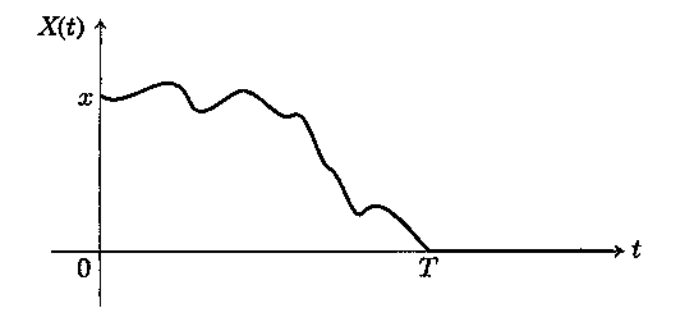


FIGURE 13.2

Définition 13.5.2. Dans la définition du mouvement brownien donné dans le paragraphe 2, remplaçons la condition (1) $X(0) = 0$ par (1′) $X(0) = x > 0$. On obtient le *mouvement brownien commencé en* x. Désignons par T le premier instant où ce mouvement brownien atteint la valeur 0 et introduisons le processus $\{Z(t) : t \geq 0\}$ défini par

$$Z(t) := \begin{cases} X(t), & \text{si } t \leq T; \\ 0, & \text{si } t > T. \end{cases}$$

Ce processus est appelé *mouvement brownien absorbé* (à l'origine) (*cf.* Figure 13.2).

13.5.3 Le mouvement brownien à dérive (« drift »)

Définition 13.5.3. Soient $\{X(t) : t \geq 0\}$ un mouvement brownien et μ un nombre réel. Le processus $\{Y(t) : t \geq 0\}$, où $Y(t) := X(t) + \mu t$, est appelé *mouvement brownien à dérive*; la constante μ est le *paramètre de dérive.*

Si $\mu \neq 0$, le processus n'est plus symétrique et le principe de réflexion ne s'applique plus pour le calcul de la loi du maximum. On a alors recours à la théorie des martingales (*cf.* § 13.6).

Proposition 13.5.2. *Le mouvement brownien avec dérive a les propriétés suivantes :*

(a) $Y(0) = 0$;

(b) *le processus est à accroissements indépendants et est stationnaire ;*

(a) *pour tout* $t > 0$*, la variable aléatoire* $Y(t)$ *suit la loi normale de paramètre* $(\mu t, \sqrt{t})$.

13.5.4 Le processus d'Ornstein-Uhlenbeck

Définition 13.5.4. Soient $\{X(t) : t \geq 0\}$ un mouvement brownien et α un nombre strictement positif. Le processus $\{Y(t) : t \geq 0\}$, où $Y(t) := e^{-(\alpha t)/2} X(e^{\alpha t})$, est appelé *processus d'Ornstein-Uhlenbeck*. Il a été proposé pour décrire la vitesse d'une particule immergée dans un fluide.

Proposition 13.5.3. *Pour tout* $t \geq 0$*, on a* $\mathbb{E}[Y(t)] = 0$ *et pour* $0 < s < t < +\infty$ *on a :* $\operatorname{Cov}(Y(s), Y(t)) = e^{-\alpha(t-s)/2}$.

Démonstration. La première relation est évidente. Ensuite,

$$\begin{aligned}\operatorname{Cov}(Y(s), Y(t)) &= \mathbb{E}[Y(s)Y(t)] = e^{-\alpha s/2} e^{-\alpha t/2}\, \mathbb{E}[X(e^{\alpha s})X(e^{\alpha t})] \\ &= e^{-\alpha s/2} e^{-\alpha t/2} e^{\alpha s} = e^{-\alpha(t-s)/2}. \quad \square\end{aligned}$$

13.5.5 Le mouvement brownien géométrique

Définition 13.5.5. Soit $\{X(t) : t \geq 0\}$ un mouvement brownien. Le processus $\{Y(t) : t \geq 0\}$, où $Y(t) := e^{X(t)}$, est appelé *mouvement brownien géométrique*.

On l'étudie surtout lorsque le mouvement brownien sous-jacent est à dérive, ou même s'il est de la forme $X(t) = u\xi(t) + \mu t$, où $\{\xi(t) : t \geq 0\}$ est le mouvemnt brownien standard. Dans la proposition suivante, on suppose cette dernière condition satisfaite.

Proposition 13.5.4. *Soient* $\{\xi(t) : t \geq 0\}$ *le mouvement brownien standard et* $\{Y(t) : t \geq 0\}$ *le processus géométrique défini par :* $Y(t) := e^{u\xi(t)+\mu t}$. *Alors* $\mathbb{E}[Y(t)] = e^{t(\mu+u^2/2)}$ *et* $\operatorname{Var} Y(t) = e^{2t(\mu+u^2/2)}(e^{tu^2} - 1)$.

Démonstration. En effet, $\mathbb{E}[Y(t)] = e^{\mu t}\,\mathbb{E}[e^{u\xi(t)}] = e^{\mu t} g_{\xi(t)}(u)$, où $g_{\xi(t)}(u)$ est la fonction génératrice des moments de $\xi(t)$, qui vaut $e^{u^2 t/2}$. Ensuite, $\mathbb{E}[Y^2(t)] = \mathbb{E}[e^{2(\mu t + u\xi(t))}] = e^{(2\mu)t} g_{\xi(t)}(2u) = e^{(2\mu)t} e^{(2u)^2 t/2} = e^{2t(\mu+u^2)}$. D'où $\operatorname{Var} Y(t) = \mathbb{E}[Y^2(t)] - (\mathbb{E}[Y(t)])^2 = e^{2t(\mu+u^2/2)}(e^{tu^2} - 1)$. $\square$

13.6 MOUVEMENT BROWNIEN ET MARTINGALES

La théorie des martingales à temps continu n'a pas été développée dans cet ouvrage. Nous allons nous contenter de donner les définitions et de décrire les analogies avec les résultats établis dans les chapitres 7–11 pour les martingales à temps discret.

Soit $\{X(t) : t \geq 0\}$ un processus défini sur un espace probabilisé $(\Omega, \mathfrak{A}, \mathrm{P})$; soit, d'autre part, $(\mathfrak{A}_t)$ $(t \geq 0)$ une famille de sous-tribus de la tribu $\mathfrak{A}$, qui est *croissante*, en ce sens que l'on a la propriété :

$$\forall\, s, t, \quad 0 \leq s < t \Longrightarrow \mathfrak{A}_s \subset \mathfrak{A}_t.$$

(1) Le processus $\{X(t) : t \geq 0\}$ est dit *adapté* à la famille $(\mathfrak{A}_t)$ $(t \geq 0)$, si, pour tout $t > 0$, la variable $X(t)$ est $\mathfrak{A}_t$-mesurable ; ce qui signifie que l'image réciproque $(X(t))^{-1}(B)$ de tout borélien B est un élément de la tribu $\mathfrak{A}_t$.

(2) Une variable aléatoire T définie sur $(\Omega, \mathfrak{A}, \mathrm{P})$, à valeurs dans $[0, +\infty]$, est appelé *temps d'arrêt relativement à* la famille $(\mathfrak{A}_t)$ $(t \geq 0)$, si, pour tout $t \geq 0$, l'évènement $\{T \leq t\}$ appartient à $\mathfrak{A}_t$.

(3) Le processus $\{X(t) : t \geq 0\}$ est dit être une *martingale par rapport à* la famille $(\mathfrak{A}_t)$ $(t \geq 0)$, si

(a) pour tout $t \geq 0$, on a : $\mathbb{E}[\,|X(t)|\,] < +\infty$;

(b) pour $0 \leq s < t$, on a : $\mathbb{E}[\,X(t)\,|\,\mathfrak{A}_s] = X(s)$.

Cette condition (b) utilise la notion d'*espérance conditionnelle* $\mathbb{E}[\,X(t)\,|\,\mathfrak{A}_s]$ de $X(t)$ par rapport à la tribu $\mathfrak{A}_s$. Ceci veut dire que $\mathbb{E}[\,X(t)\,|\,\mathfrak{A}_s]$ est une variable aléatoire $\mathfrak{A}_s$-mesurable et qui est telle que pour tout $A \in \mathfrak{A}_s$ on a l'identité : $\mathbb{E}[I_A X(t)] = \mathbb{E}[I_A \mathbb{E}[\,X(t)\,|\,\mathfrak{A}_s]\,]$.

Par conséquent, la condition (b) s'exprime encore en disant que pour tout $A \in \mathfrak{A}_s$, la relation $\mathbb{E}[I_A X(t)] = \mathbb{E}[I_A X(s)]$ est satisfaite (en particulier, $\mathbb{E}[X(t)] = \mathbb{E}[X(s)]$). Cette relation est à rapprocher de la condition (b), Définition 7.1.2, pour les martingales (X_n) à temps discret, qui s'exprimait par $\mathbb{E}[I_A X_n] = \mathbb{E}[I_A X_{n+1}]$ pour tout A appartenant à la tribu engendrée par le vecteur $(Y_0, Y_1, \ldots, Y_n)$, une tribu qui joue le rôle de $\mathfrak{A}_s$. Notons encore que la condition (b) implique que la martingale $\{X(t) : t \geq 0\}$ est *adaptée à* la famille $(\mathfrak{A}_t)$ $(t \geq 0)$.

Théorème 13.6.1. *Le mouvement brownien standard* $\{X(t) : t \geq 0\}$ *est une martingale par rapport à la famille* $(\mathfrak{A}_t)$ $(t \geq 0)$*, où* $\mathfrak{A}_t$ *est la tribu engendrée par les variables* X_s $(0 \leq s \leq t)$.

Démonstration. Soit $0 \leq s < t < +\infty$. Alors $\mathbb{E}[X(t)\,|\,\mathfrak{A}_s] = \mathbb{E}[X(s) + (X(t) - X(s))\,|\,\mathfrak{A}_s] = \mathbb{E}[X(s)\,|\,\mathfrak{A}_s] + \mathbb{E}[X(t) - X(s)\,|\,\mathfrak{A}_s]$. Or, $\mathbb{E}[X(s)\,|\,\mathfrak{A}_s] = X(s)$, en appliquant l'analogue de la Règle 2, § 1.13. Puis, $\mathbb{E}[X(t) - X(s)\,|\,\mathfrak{A}_s] = \mathbb{E}[X(t) - X(s)]$ en appliquant l'analogue de la Règle 1, § 1.13, puisque $X(t) - X(s)$ est indépendante de la tribu $\mathfrak{A}_s$, le processus étant à accroissements indépendants. Enfin, $\mathbb{E}[X(t) - X(s)] = 0$, puisque les variables $X(s)$ et $X(t)$ ont même espérance mathématique. D'où l'identité $\mathbb{E}[X(t)\,|\,\mathfrak{A}_s] = X(s)$. □

Les inégalités maximales pour les martingales à temps discret ont été démontrées dans le chapitre 10. Nous donnons, sans démonstration, leur énoncé pour les martingales à temps continu. Le Théorème 13.6.1 permet alors d'utiliser ces inégalités pour le mouvement brownien standard.

Inégalités maximales.

$$\forall\, \varepsilon > 0 \quad \mathrm{P}\Big\{\sup_{0\le s\le t} |X(s)| \ge \varepsilon\Big\} \le \frac{\mathbb{E}[\,|X(t)|\,]}{\varepsilon}; \tag{13.6.1}$$

$$\forall\, \varepsilon > 0 \quad \mathrm{P}\Big\{\sup_{0\le s\le t} |X(s)| \ge \varepsilon\Big\} \le \frac{\mathbb{E}[\,X^2(t)\,]}{\varepsilon^2} = \frac{t}{\varepsilon^2}. \tag{13.6.2}$$

Théorème 13.6.2. *Le processus $U(t) = X^2(t) - t$ est une martingale par rapport à la famille $(\mathfrak{A}_t)$ $(t \ge 0)$, où $\mathfrak{A}_t$ est la tribu $\mathfrak{T}((X_s)\ (0 \le s \le t))$ engendrée par la famille des variables aléatoires (X_s) $(0 \le s \le t)$.*

Démonstration. Soient $0 \le s < t < +\infty$. Alors $\mathbb{E}[U(t)\,|\,\mathfrak{A}_s] = \mathbb{E}[X^2(t)\,|\,\mathfrak{A}_s] - t$. Or, $\mathbb{E}[X^2(t)\,|\,\mathfrak{A}_s] = \mathbb{E}[(X(s) + (X(t) - X(s)))^2\,|\,\mathfrak{A}_s] = \mathbb{E}[X^2(s)\,|\,\mathfrak{A}_s] + 2\,\mathbb{E}[X(s)(X(t)-X(s))\,|\,\mathfrak{A}_s] + \mathbb{E}[(X(t)-X(s))^2\,|\,\mathfrak{A}_s] = X^2(s) + 2\,X(s)\,\mathbb{E}[X(t)-X(s)\,|\,\mathfrak{A}_s] + \mathbb{E}[(X(t)-X(s))^2\,|\,\mathfrak{A}_s] = X^2(s) + 2\,X(s)\,\mathbb{E}[X(t)-X(s)] + \mathbb{E}[(X(t)-X(s))^2] = X^2(s) + t - s$ et par conséquent $\mathbb{E}[U(t)\,|\,\mathfrak{A}_s] = X^2(s) + (t-s) - t = X^2(s) - s = U(s)$. □

La *martingale de Wald* associée à une suite de variables aléatoires, indépendantes, identiquement distribuées, admettant une fonction génératrice des moments a été construite au paragraphe 7.3. Le théorème suivant montre que l'on peut construire une martingale analogue, qui est, elle, associée au mouvement brownien. Par analogie avec le cas discret, on introduit le processus $\{V(t) : t \ge 0\}$, où

$$V(t) := \frac{e^{uX(t)}}{\mathbb{E}[e^{uX(t)}]} = \frac{e^{uX(t)}}{e^{u^2t/2}} = \exp\Big(u\Big(X(t) - \frac{ut}{2}\Big)\Big).$$

Théorème 13.6.3. *Le processus $\{V(t) : t \ge 0\}$, est une martingale par rapport à la famille $(\mathfrak{A}_t)$ $(t \ge 0)$, où $\mathfrak{A}_t := \mathfrak{T}((X_s)\ (0 \le s \le t))$.*

Démonstration. Soient $0 \le s < t < +\infty$. Alors

$$\begin{aligned}
\mathbb{E}[V(t)\,|\,\mathfrak{A}_s] &= e^{-u^2t/2}\,\mathbb{E}[e^{uX(t)}\,|\,\mathfrak{A}_s] = e^{-u^2t/2}\,\mathbb{E}[e^{u(X(s)+(X(t)-X(s))}\,|\,\mathfrak{A}_s] \\
&= e^{-u^2t/2}\,e^{uX(s)}\,\mathbb{E}[e^{u(X(t)-X(s))}\,|\,\mathfrak{A}_s] \\
&= e^{-u^2t/2}\,e^{uX(s)}\,\mathbb{E}[e^{u(X(t)-X(s))}] = e^{-u^2t/2}\,e^{uX(s)}\,e^{u^2(t-s)/2} \\
&= e^{-u^2s/2}\,e^{uX(s)} = V(s). \quad \square
\end{aligned}$$

Remarque. Pour le mouvement brownien, de dérive μ et de coefficient de diffusion σ^2, la martingale associée est définie par : $e^{uX(t)-ut\mu-u^2\sigma^2t/2}$.

Application. La martingale $\{V(t) : t \geq 0\}$, définie dans le Théorème 13.6.3, est *à valeurs positives.* En utilisant l'inégalité maximale (13.6.1) et en notant que l'on a $\mathbb{E}[V(t)] = 1$, on obtient :

$$\forall\, \varepsilon > 0 \quad \mathrm{P}\Big\{\sup_{0\leq s\leq t} V(s) \geq \varepsilon\Big\} \leq \frac{\mathbb{E}[V(t)]}{\varepsilon} = \frac{1}{\varepsilon}.$$

Cette dernière inégalité permet de déduire des inégalités maximales pour le mouvement brownien *à dérive.* En effet, pour $u \geq 0$, on a identité entre les évènements :

$$\Big\{\sup_{0\leq s\leq t}\Big(X(s) - \frac{u}{2}s\Big) \geq v\Big\} \quad \text{et} \quad \Big\{\sup_{0\leq s\leq t} V(s) \geq e^{uv}\Big\}.$$

Comme $\mathrm{P}\Big\{\sup_{0\leq s\leq t} V(s) \geq e^{uv}\Big\} \leq e^{-uv}$, on en tire, pour $u, v \geq 0$, l'inégalité

$$\mathrm{P}\Big\{\sup_{0\leq s\leq t}\Big(X(s) - \frac{u}{2}s\Big) \geq v\Big\} \leq e^{-uv}.$$

Cette inégalité, relative au processus à dérive $\{X(t) - ut/2 : t \geq 0\}$, est utilisée dans les problèmes de ruine.

13.7 TEMPS D'ATTEINTE

Considérons un mouvement brownien $\{X(t) : t \geq 0\}$, où $X(t)$ est une variable aléatoire normale d'espérance μt et d'écart-type $\sigma\sqrt{t}$. On suppose, de plus, $X(0) = 0$. La fonction génératrice des moments est donnée par $g(u) = \mathbb{E}[e^{uX(t)}] = e^{\mu t u + \sigma^2 t u^2/2}$ ($t \geq 0$, $u \in \mathbb{R}$). Pour $a, b > 0$, on définit le *temps d'atteinte* T par :

$$T := \inf\{t \geq 0 : X(t) = -a \text{ ou } X(t) = +b\}.$$

Propriété 13.7.1. *Le temps d'atteinte T est un temps d'arrêt adapté à $\{X(t) : t \geq 0\}$.*

Propriété 13.7.2. *On a :* $\mathrm{P}\{T < +\infty\} = 1$.

Démonstration. En effet, pour tout $t > 0$ fini, on a les majorations :

$$\mathrm{P}\{T = +\infty\} \leq \mathrm{P}\{T > t\} = \mathrm{P}\{-a < X(t) < b\} = \int_{-a}^{+b} f_{X(t)}(x)\, dx \leq \frac{a+b}{\sigma\sqrt{2\pi t}}.$$

D'où $\mathrm{P}\{T = +\infty\} = 0$, en faisant tendre t vers l'infini. □

Ci-après, on utilise l'analogue continu du théorème d'arrêt pour les martingales dominées, à savoir le Théorème 8.3.1. Les hypothèses sont : (a) $\mathrm{P}\{T < +\infty\} = 1$ (c'est le contenu de la Propriété 13.7.2) ; (b) le processus arrêté $\{X(T \wedge t) : t \geq 0\}$ est borné (ce qui est banal ici). Lorsque le processus $\{X(t) : t \geq 0\}$ est une martingale (c'est le cas lorsque $\mu = 0$, voir

ci-dessous), on peut conclure que $\mathbb{E}[X(T)] = \mathbb{E}[X(0)]$. On utilise aussi les notations $r_a := \mathrm{P}\{X(T) = -a\}$ et $r_b := \mathrm{P}\{X(T) = b\}$, de sorte que, d'après la Propriété 13.7.2, on a : $r_a + r_b = 1$.

13.7.1 Cas symétrique $\mu = 0$

Dans ce cas, le mouvement brownien $\{X(t) : t \geq 0\}$ est une martingale, d'après le Théorème 13.6.1. On peut alors appliquer le théorème d'arrêt et écrire : $\mathbb{E}[X(T)] = \mathbb{E}[X(0)]$. Comme la loi de $X(T)$ est la loi de Bernoulli $r_a\varepsilon_{-a} + (1-r_a)\varepsilon_{+b}$, on en déduit $0 = \mathbb{E}[X(0)] = \mathbb{E}[X(T)] = r_a(-a) + (1-r_a)b$, d'où

$$r_a = \frac{b}{a+b}, \tag{13.7.1}$$

une expression qui est indépendante de σ^2.

13.7.2 Cas non-symétrique $\mu \neq 0$

Dans ce cas, le processus $\{X(t) : t \geq 0\}$ n'est plus une martingale. On lui associe alors la martingale de Wald $\{Z(t) : t \geq 0\}$, définie, pour tout $u \neq 0$, par

$$Z(t) := \frac{e^{uX(t)}}{g(u)} = \exp\Big(uX(t) - ut\Big(\mu + \frac{u\sigma^2}{2}\Big)\Big).$$

Choisissons $u = -2\mu/\sigma^2$ de sorte que l'on ait $\mu + u\sigma^2/2 = 0$ et considèrons ainsi la martingale

$$Z(t) = \exp\Big(-\frac{2\mu}{\sigma^2}X(t)\Big) \quad (t \geq 0).$$

La variable $Z(T) = \exp\Big(-\frac{2\mu}{\sigma^2}X(T)\Big)$ ne dépend alors de T que par l'intermédiaire de $X(T)$. On est dans les conditions d'applications du théorème d'arrêt : $\mathbb{E}[Z(T)] = \mathbb{E}[Z(0)] = 1$. Par conséquent, $r_a\, e^{-2\mu(-a)/\sigma^2} + (1-r_a)\, e^{-2\mu b/\sigma^2)} = 1$, soit

$$r_a = \frac{1 - e^{-2\mu b/\sigma^2}}{e^{2\mu a/\sigma^2} - e^{-2\mu b/\sigma^2}} = \frac{1 - e^{2\mu b/\sigma^2}}{1 - e^{2\mu(a+b)/\sigma^2}}. \tag{13.7.2}$$

Remarque. Lorsque μ/σ^2 tend vers 0, on a $r_a \to b/(a+b)$, l'expression trouvée dans le cas *symétrique*. Ceci est, en particulier, le cas si σ^2 est fixe et μ tend vers 0, ou bien, lorsque μ est fixe et σ^2 tend vers $+\infty$.

13.7.3 Mouvement brownien débutant en y

Pour y réel et $t \geq 0$, considérons le mouvement brownien débutant en y défini par $S^y(t) := y + X(t)$. Posons $T_c^y := \inf\{t \geq 0 : S^y(t) = c\}$ et pour $c < y < d$ évaluons la probabilité $\mathrm{P}\{T_c^y < T_d^y\}$ pour que le processus, partant de y, atteigne c *avant* d.

Étant donné $c < y < d$, posons $a := y - c$ et $b := d - y$. Les formules (13.7.1) et (13.7.2) donnent :

$$\mathrm{P}\{T_c^y < T_d^y\} = \begin{cases} \dfrac{d-y}{d-c}, & \text{si } \mu = 0\,; \\ \dfrac{1 - e^{2\mu(d-y)/\sigma^2}}{1 - e^{2\mu(d-c)/\sigma^2}}, & \text{si } \mu \neq 0. \end{cases} \tag{13.7.3}$$

Pour calculer $\mathrm{P}\{T_d^y < T_c^y\}$, il suffit de faire les substitutions $\mu \leftarrow -\mu$ et $d - y \leftarrow y - c$. On obtient alors :

$$\mathrm{P}\{T_d^y < T_c^y\} = \begin{cases} \dfrac{y-c}{d-c}, & \text{si } \mu = 0\,; \\ \dfrac{1 - e^{-2\mu(y-c)/\sigma^2}}{1 - e^{-2\mu(d-c)/\sigma^2}}, & \text{si } \mu \neq 0. \end{cases} \tag{13.7.4}$$

Dans les deux cas $\mu = 0$ et $\mu \neq 0$, on constate que l'on a :

$$\mathrm{P}\{T_c^y < T_d^y\} + \mathrm{P}\{T_d^y < T_c^y\} = 1.$$

Considérons l'évènement « le processus $\{S^y(t) : t \geq 0\}$ atteint c au moins une fois », un évènement qui s'exprime encore par $\{T_c^y < +\infty\}$. Pour tout y réel, la famille des variables aléatoires (T_d^y) $(d > y)$ est croissante et tend vers l'infini lorsque d tend vers l'infini. Il en résulte que la famille des évènements $\{T_c^y < T_d^y\}$ $(d > y)$ est *croissante*, lorsque d tend vers l'infini et sa limite est $\{T_c^y < +\infty\}$. On a donc : $\mathrm{P}\{T_c^y < +\infty\} = \lim_{d \to +\infty} \mathrm{P}\{T_c^y < T_d^y\}$. En vertu de (13.7.3), on en déduit :

$$\mathrm{P}\{T_c^y < +\infty\} = \begin{cases} 1, & \text{si } \mu = 0\,; \\ e^{-2\mu(y-c)/\sigma^2} < 1, & \text{si } \mu > 0\,; \\ 1, & \text{si } \mu < 0. \end{cases} \tag{13.7.5}$$

De même, $\mathrm{P}\{T_d^y < +\infty\} = \lim_{c \to -\infty} \mathrm{P}\{T_d^y < T_c^y\}$ et de (7.4) on tire :

$$\mathrm{P}\{T_d^y < +\infty\} = \begin{cases} 1, & \text{si } \mu = 0\,; \\ 1, & \text{si } \mu > 0\,; \\ e^{-2|\mu|(d-y)/\sigma^2} < 1, & \text{si } \mu < 0. \end{cases} \tag{13.7.6}$$

Ce résultat montre que le mouvement brownien *symétrique* $(\mu = 0)$ est *récurrent*, alors que celui *avec dérive non-nulle* $(\mu \neq 0)$ est *transient*.

13.8 LOI DE PROBABILITÉ DU TEMPS D'ATTEINTE

Soit $\{X(t) : t \geq 0\}$ un mouvement brownien tel que $X(t)$ soit d'espérance μt et de variance $\sigma^2 t$ avec $\sigma > 0$. Dans le paragraphe 13.3.1, ont été données, en (13.3.2) et (13.3.3), la fonction de répartition et la densité du temps d'atteinte $T_x := T_x^0$ en un point x, pour un mouvement brownien *standard*,

c'est-à-dire tel que $\mu = 0$ et $\sigma = 1$, débutant en 0. Lorsque μ est toujours nulle et que σ n'est plus nécessairement égal à 1, ces formules deviennent :

$$\mathrm{P}\{T_x^0 \le t\} = \frac{2}{\sqrt{2\pi}} \int_{|x|/(\sigma\sqrt{t})}^{+\infty} e^{-y^2/2}\, dy \quad (t > 0); \tag{13.8.1}$$

$$f_{T_x^0}(t) = \frac{1}{\sqrt{2\pi}} \frac{|x|}{\sigma}\, t^{-3/2}\, e^{-x^2/(2\sigma^2 t)} \quad (t > 0). \tag{13.8.2}$$

On déduit de (13.8.1) que, pour $x \neq 0$, on a : $\lim_{t\to+\infty} \mathrm{P}\{T_x^0 > t\} = 0$ et $\lim_{t\to 0+} \mathrm{P}\{T_x^0 > t\} = 1$. Par conséquent, T_x^0 est presque sûrement fini et appartient à l'intervalle $]0, +\infty[$, presque sûrement.

Proposition 13.8.1. *Si* $\mu = 0$, $\sigma \neq 0$ *et* $x \neq 0$, *on a :* $\mathbb{E}[T_x^0] = +\infty$.

Démonstration. Dans l'intervalle $[0, +\infty]$, on a

$$\mathbb{E}[T_x^0] = \int_0^{+\infty} \mathrm{P}\{T_x^0 > t\}\, dt,$$

soit, en posant $\alpha := 2/\sqrt{2\pi}$, $\beta := |x|/\sigma$,

$$\begin{aligned}
\mathbb{E}[T_x^0] &= \alpha \int_0^{+\infty} dt \int_0^{\beta/\sqrt{t}} e^{-u^2/2}\, du \\
&= \alpha \int_0^{+\infty} e^{-u^2/2}\, du \int_0^{\beta^2/u^2} dt \\
&= \alpha\, \beta^2 \int_0^{+\infty} \frac{1}{u^2} e^{-u^2/2}\, du = +\infty. \quad \square
\end{aligned}$$

Considérons maintenant le cas avec dérive μ non nulle. Dans les formules (13.7.5) et (13.7.6), prenons $c < y = 0 < d$. On obtient $\mathrm{P}\{T_c^0 < +\infty\} < 1$, si $\mu > 0$ et donc si $\mu c < 0$. De même, $\mathrm{P}\{T_d^0 < +\infty\} < 1$, si $\mu < 0$ et donc si $\mu d < 0$. On obtient également $\mathrm{P}\{T_c^0 < +\infty\} = 1$, si $\mu < 0$ et donc $\mu c > 0$, puis $\mathrm{P}\{T_d^0 < +\infty\} = 1$, si $\mu > 0$ et donc $\mu d > 0$. On peut donc récrire :

si $\mu x < 0$, alors $\mathrm{P}\{T_x^0 < +\infty\} < 1$;

si $\mu x > 0$, alors $\mathrm{P}\{T_x^0 < +\infty\} = 1$.

Pour le calcul de la loi de probabilité de T_x^0, nous faisons appel de nouveau à la *martingale de Wald* (*cf.* § 13.7.2)[1]

$$Z(t) = \exp\Big(uX(t) - ut\Big(\mu + \frac{u\sigma^2}{2}\Big)\Big).$$

Prenons $x > 0$ et distinguons les cas $\mu \ge 0$ et $\mu < 0$.

(a) $\mu \ge 0$; pour $u \ge 0$ et $0 \le t \le T_x^0$, on a $0 \le Z(t) \le \exp(ux)$. Le processus arrêté (en T_x^0) est donc borné. Comme $\mathrm{P}\{T_x^0 < +\infty\} = 1$, on peut appliquer l'analogue continu du théorème d'arrêt pour les martingales dominées

[1] Nous remercions M. Émery de nous avoir explicité cette méthode de calcul.

(*cf.* Théorème 8.3.1) et écrire :

$$1 = \mathbb{E}[Z(0)] = \mathbb{E}[Z(T_x^0)] = \mathbb{E}\Big[\exp\Big(ux - uT_x^0\Big(\mu + \frac{u\sigma^2}{2}\Big)\Big)\Big].$$

En prenant pour u la solution positive $u = (\sqrt{\mu^2 + 2s\sigma^2} - \mu)/\sigma^2$ de l'équation $u\mu + u^2\sigma^2/2 = s$, on trouve, pour tout $s \geq 0$, l'identité

$$\mathbb{E}[e^{-sT_x^0}] = e^{-(\sqrt{\mu^2+2s\sigma^2}-\mu)(x/\sigma^2)}. \tag{13.8.3}$$

On obtient ainsi la *transformée de Laplace* de la loi de T_x^0. La transformée de Laplace *inverse* du membre de droite de (13.8.3) peut être obtenue (voir Exercice 8). On obtient, pour T_x^0, la densité de probabilité :

$$f_{T_x^0}(t) = \frac{1}{\sqrt{2\pi}}\frac{x}{\sigma}t^{-3/2}\exp\Big(-\frac{1}{2\sigma^2}\Big(\frac{x}{\sqrt{t}} - \mu\sqrt{t}\Big)^2\Big) \quad (t > 0). \tag{13.8.4}$$

Posant $f(s) := \mathbb{E}[e^{-sT_x^0}]$, on trouve directement l'espérance mathématique de T_x^0, à savoir $\mathbb{E}[T_x^0] = -f'(0) = x/\mu$.

(b) $\mu < 0$; pour tout $u \geq -2\mu/\sigma^2$ et $0 \leq t \leq T_x^0$, on a : $0 \leq Z(t) \leq \exp(ux)$. D'autre part, comme $Z(\infty) = 0$, on en déduit :

$$1 = \mathbb{E}[Z(T_x^0)] = \mathbb{E}\Big[\exp\Big(ux - uT_x^0\Big(\mu + \frac{u\sigma^2}{2}\Big)\Big)\, I_{\{T_x^0<\infty\}}\Big].$$

Pour tout $s \geq 0$, on trouve la transformée de Laplace

$$\mathbb{E}[e^{-sT_x^0}\, I_{\{T_x^0<\infty\}}] = e^{-(\sqrt{\mu^2+2s\sigma^2}-\mu)(x/\sigma^2)}. \tag{13.8.5}$$

En particulier, pour $s = 0$, on obtient l'évaluation :

$$\mathrm{P}\{T_x^0 < +\infty\} = e^{2\mu x/\sigma^2},$$

une expression strictement inférieure à 1, puisque $\mu < 0$.

13.9 COMPLÉMENTS ET EXERCICES

13.1 Soit $\{X(t) : t \geq 0\}$ un mouvement brownien standard.

(a) Pour $0 < s < t$, on a $\mathrm{Cov}(X(s), X(t)) = s$, soit, de façon générale, $\mathrm{Cov}(X(s), X(t)) = \min(s, t)$.

(b) Pour $0 < s < t$ la loi de $X(s)$, conditionnellement à $\{X(t) = b\}$, est la loi normale $N(\mu, \sigma)$ avec $\mu = bs/t$ et $\sigma = \sqrt{s(1 - s/t)}$.

(c) Pour $0 < s < t < 1$, on a : $\mathrm{Cov}((X(s), X(t)) \,|\, X(1) = 0) = s(1 - t)$. [Utiliser la formule (1.13.4), avec $Z \leftarrow X(1)$, $g \leftarrow Id$, $X \leftarrow X(s)X(t)$, $Y \leftarrow X(t)$ et (b).]

13.2 Pour $a > 0$, calculer $\mathrm{P}\{\inf_{0\leq s\leq t} X(s) \leq 0 \,|\, X(0) = a\}$.

13.3 Soient $0 < t_0 < t_1$ et $a \in \mathbb{R}$. Si $X(t_0) = a$, la probabilité $P(a)$ pour que $X(t)$ admette au moins un zéro dans $[t_0, t_1]$ est donnée par :

$$P(a) = \frac{|a|}{\sqrt{2\pi}} \int_0^{t_1 - t_0} u^{-3/2} e^{-a^2/(2u)}\, du.$$

13.4 Soient $0 < t_0 < t_1$; la probabilité α pour que, si $X(0) = 0$, la fonction $X(t)$ admette au moins un zéro dans $[t_0, t_1]$ est égale à :

$$\alpha = \frac{2}{\pi} \operatorname{Arc\,cos} \sqrt{\frac{t_0}{t_1}}.$$

[Soit B l'évènement « $X(t)$ admet au moins un zéro dans $[t_0, t_1]$ ». Calculer $\mathrm{P}(B)$ par la formule (*cf.* formule (1.13.3))

$$\mathrm{P}(B) = \int_{\mathbb{R}} \mathrm{P}\{B \,|\, |X(t_0)| = a\}\, f_{|X(t_0)|}(a)\, da,$$

où $f_{|X(t_0)|}(a)$ est la densité de $|X(t_0)|$ et utiliser l'Exercice 13.3.]

13.5 Deux mouvements browniens *indépendants* $\{X_1(t) : t \geq 0\}$ et $\{X_2(t) : t \geq 0\}$, tels que $X_i(t)$ est d'espérance $\mu_i t$ et de variance $\sigma_i^2 t$ ($i = 1, 2$), se trouvent à la date $t = 0$ dans les états x_1 et x_2 ($x_1 < x_2$), respectivement. Calculer la probabilité pour que ces deux processus ne se rencontrent jamais. [Distinguer les cas $\mu_1 - \mu_2 = 0$, > 0 et < 0.]

13.6 Soit $\{X(t) : t \geq 0\}$ un mouvement brownien tel que $X(0) = 0$ et tel que $X(t)$ soit d'espérance μt et de variance $\sigma^2 t$ avec $\sigma > 0$. On pose $Y := -\inf_{t \geq 0} X(t)$ et $Z := \sup_{t \geq 0} X(t)$. Si $\mu > 0$ (resp. si $\mu < 0$), alors Y (resp. Z) est une variable aléatoire exponentielle.

13.7 Soient n et p deux entiers tels que $n \geq 2$, $1 \leq p \leq n - 1$ et ${}^tX := (X_1, \ldots, X_n)$ un vecteur aléatoire suivant une loi normale $\mathcal{N}(0, \Gamma)$, avec Γ *régulière*. On pose : ${}^tY := (X_1, \ldots, X_p)$ et ${}^tZ := (X_{p+1}, \ldots, X_n)$. On peut ainsi partitionner la matrice Γ des variances-covariances comme $\Gamma = \begin{pmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22} \end{pmatrix}$, avec $\Gamma_{11} := \mathbb{E}[Y\, {}^tY]$, $\Gamma_{22} := \mathbb{E}[Z\, {}^tZ]$, $\Gamma_{12} = {}^t\Gamma_{21} := \mathbb{E}[Y\, {}^tZ]$. On se propose de déterminer la loi de Y conditionnelle à $\{Z = z\}$, où z est un certain élément de $\mathbb{R}^{n-p}$.

(a) On pose : $Y' := Y - A\,Z$, $Z' := Z$, où A est une matrice $(p \times (n - p))$. Déterminer A pour que les vecteurs Y' et Z' soient non corrélés. On trouve : $A = \Gamma_{12}\, \Gamma_{22}^{-1}$.

(b) Cette matrice A étant fixée, les vecteurs Y' et Z' sont indépendants et Y' suit une loi normale $\mathcal{N}(0,\Gamma')$ avec $\Gamma' := \Gamma_{11} - \Gamma_{12}\,\Gamma_{22}^{-1}\,\Gamma_{21}$.

(c) Le vecteur $Y = \Gamma_{22}^{-1}\,\Gamma_{21}\,Z + Y'$ est la somme de deux vecteurs indépendants. La loi de Y, conditionnellement à $\{Z = z\}$, est la loi normale $\mathcal{N}(\Gamma_{12}\Gamma_{22}^{-1}z, \Gamma')$. [Appliquer la formule (1.13.8).]

13.8 Pour $t > 0$, on pose

$$I(t) = \int_0^{+\infty} \exp\Bigl(-\frac{1}{2}\Bigl(u - \frac{t}{u}\Bigr)^2\Bigr)\,du; \tag{α}$$

$$J(t) = \int_0^{+\infty} \exp\Bigl(-\frac{1}{2}\Bigl(u^2 + \frac{t^2}{u^2}\Bigr)\Bigr)\,du. \tag{β}$$

(a) On a les évaluations : $I(t) = \sqrt{2\pi}/2$; $J(t) = (\sqrt{2\pi}/2)\,e^{-t}$.

(b) Pour $s \ge 0$ et $x > 0$, on a :

$$\int_0^{+\infty} e^{-st}\,\frac{x}{2\sqrt{\pi}}\,t^{-3/2}\,e^{-x^2/(4t)}\,dt = e^{-x\sqrt{s}}.$$

[Faire le changement de variables $t = x^2/(2u^2)$ et utiliser (a).]

(c) La transformée de Laplace de la fonction $f(t)$ est définie par $\mathcal{L}\,f(t) := \int_0^{+\infty} e^{-st}\,f(t)\,dt := g(s)$. On écrit $f(t) := \mathcal{L}^{-1}\,g(s)$ pour la transformée de Laplace inverse. Les relations suivantes sont immédiates :

$\mathcal{L}^{-1}\,a\,g(s) = a\,\mathcal{L}^{-1}\,g(s) = a\,f(t)$;
$\mathcal{L}^{-1}\,g(s-a) = e^{at}\,\mathcal{L}^{-1}\,g(s) = e^{at}f(t)$;
$\mathcal{L}^{-1}\,g(s/a) = a\,f(at)$.

On déduit de ces relations et de (b) que l'on a :

$$\mathcal{L}^{-1}\,e^{-(\sqrt{\mu^2+2s\sigma^2}-\mu)(x/\sigma^2)} = f_{T_x^0}(t),$$

où $f_{T_x^0}(t)$ est donnée par (13.8.4).

13.9 On reprend toutes les notations du Théorème 13.4.2 et on utilise le fait que les vecteurs Y et Z sont *indépendants*, comme démontré au début de la démonstration de ce théorème. Considérons une suite de n couples de nombres réels $(a_1, b_1), \ldots, (a_n, b_n)$, tels que $a_1 < b_1, \ldots, a_n < b_n$. Pour k assez grand, on obtient l'encadrement

$$\mathrm{P}\{\bigcap_{i=1}^{n}\{a_i + 1/k < Y_i \le b_i - 1/k\}\} \le \mathrm{P}\{\bigcap_{i=1}^{n}\{a_i < Y_i + Z_i \le b_i\} \Bigm| |X(1)| \le 1/k\}$$
$$\le \mathrm{P}\{\bigcap_{i=1}^{n}\{a_i - 1/k < Y_i \le b_i + 1/k\}\},$$

d'où l'on tire

$$\mathrm{P}\{\bigcap_{i=1}^{n}\{a_i < Y_i \le b_i\}\} = \mathrm{P}\{\bigcap_{i=1}^{n}\{a_i < X(t_i) \le b_i\} \Bigm| |X(1)| = 0\}.$$

Solutions des Exercices

CHAPITRE 2

2.1 A l'aide des variable aléatoires Y_i, l'évènement $\{T_A = n\}$ s'exprime comme $\{Y_0 \notin A, \ldots, Y_{n-1} \notin A, Y_n \in A\}$, qui appartient évidemment à $\mathfrak{T}(Y_0, \ldots, Y_n)$. De même, l'évènement $\{T_A^{(r)} = n\}$ se réalise si et seulement s'il existe un ensemble I d'entiers positifs, de cardinal r, tel que $\max I = n$ et tel que pour tout $i = 0, 1, \ldots, n$ on ait $Y_i \in A$, si $i \in I$ et $Y_i \notin A$, si $i \notin I$. Un tel évènement est aussi dans $\mathfrak{T}(Y_0, \ldots, Y_n)$.

2.2 Le dernier instant où le processus est dans A peut s'exprimer comme $T'_A := \sup\{n \geq 0 : Y_n \in A\}$. Or, $\{T'_A = n\}$ n'est autre que l'évènement $\{Y_n \in A\} \cap \bigcap_{k \geq n+1} \{Y_k \notin A\}$. Dans le cas des variables aléatoires indépendantes considérées, $\{T'_A = 1\} = \{Z_0 = -1, Z_1 = 0, Z_2 = 1\}$. Si cet évènement appartenait à $\mathfrak{T}(Z_0, Z_1) = \mathfrak{T}(Y_0, Y_1)$, il devrait être réunion d'évènements de la forme $\{Z_0 = i_0, Z_1 = i_1\}$, avec $i_0, i_1 = \pm 1$. On devrait donc avoir $p \leq p^2$ ou $p \leq pq$, d'où une contradiction si l'on suppose $0 < p < 1$.

2.3 Pour chaque $n \geq 0$ on a $\{T_A = n\} = \{Y_0 = 0, \ldots, Y_{n-1} = 0, Y_n = 1\}$, ce qui redémontre que T_A est un temps d'arrêt. De plus la probabilité de l'évènement est égale à $q^n p$. Supposons $r \geq 2$ et reprenons les mêmes notations que dans la solution de l'Exercice 2.2. Le nombre de sous-ensembles I d'entiers qui conviennent est égal à $\binom{n}{r-1}$ et la probabilité de $\bigcap_{i \in I}\{Y_i = 1\} \cap \bigcap_{i \notin I}\{Y_i = 0\}$ vaut $p^r q^{n+1-r}$, d'où $\mathrm{P}\{T_A^{(r)} = n\} = \binom{n}{r-1} p^r q^{n+1-r}$.

2.4 L'évènement $\{T = n\}$ ne se produit que si $n = 2$ ou 3. On a $\{T = 2\} = \{Y_0 = 1\} \in \mathfrak{T}(Y_0, Y_1, Y_2)$ et $\{T = 3\} = \{Y_0 = 0\} \in \mathfrak{T}(Y_0, Y_1, Y_2, Y_3)$. De là, T est un temps d'arrêt.
De même, $\{T' = n\}$ est l'évènement impossible si $n \neq 2, 3$ et $\{T' = 2\} = \{Y_3 = 1\} \in \mathfrak{T}(Y_0, Y_1, Y_2, Y_3)$. D'autre part, l'évènement $\{Y_3 = 1\}$ n'appartient pas à $\mathfrak{T}(Y_0, Y_1, Y_2)$.

2.5 Notons que T ne prend que des valeurs paires. D'abord, $\{T = 2k, Y_{2k} \leq 1\} = \{T = 2k\}$; ensuite, pour $0 \leq i \leq k-1$, on a : $\{T = 2i, Y_{2k} \leq 1\} = \{T = 2i, Z_{T+1} + \cdots + Z_{T+2(k-i)} \leq 0\}$ et de même, $\{T = 2k, Y_{2k} \geq 1\} = \{T = 2k\}$; puis, pour $0 \leq i \leq k-1$, on a : $\{T = 2i, Y_{2k} \geq 1\} = \{T = 2i, Z_{T+1} + \cdots + Z_{T+2(k-i)} \geq 0\}$. D'après le Théorème 4.2, $\mathrm{P}\{T = 2i, Z_{T+1} + \cdots + Z_{T+2(k-i)} \leq 0\} = \mathrm{P}\{T = 2i\}$ $\mathrm{P}\{Z_{T+1} + \cdots + Z_{T+2(k-i)} \leq 0\} = \mathrm{P}\{T = 2i\}\mathrm{P}\{Z_1 + \cdots + Z_{2(k-i)} \leq 0\}$, une expression encore égale à $\mathrm{P}\{T = 2i\}\mathrm{P}\{Z_1 + \cdots + Z_{2(k-i)} \geq 0\}$, puisque l'on a supposé $p = q = \frac{1}{2}$. Pour $0 \leq i \leq k$, on en déduit : $\mathrm{P}\{T = 2i, Y_{2k} \leq 1\} = \mathrm{P}\{T = 2i, Y_{2k} \geq 1\}$ et en sommant sur i, $\mathrm{P}\{T \leq 2k, Y_{2k} \leq 1\} = \mathrm{P}\{T \leq 2k, Y_{2k} \geq 1\}$. On en tire : $\mathrm{P}\{T \leq 2k\} = \mathrm{P}\{T \leq 2k, Y_{2k} \leq 1\} + \mathrm{P}\{T \leq 2k, Y_{2k} \geq 1\} - \mathrm{P}\{T \leq 2k, Y_{2k} = 1\} = 2\,\mathrm{P}\{T \leq 2k, Y_{2k} \geq 1\} - \mathrm{P}\{T \leq 2k, Y_{2k} = 1\} = 2\,\mathrm{P}\{T \leq 2k, Y_{2k} > 1\} + \mathrm{P}\{T \leq 2k, Y_{2k} = 1\} = 2\,\mathrm{P}\{Y_{2k} > 1\} + \mathrm{P}\{Y_{2k} = 1\}$, puisque chacun des évènements $\{Y_{2k} > 1\}$, $\{Y_{2k} = 1\}$ entraîne la réalisation de $\{T = 2k\}$. Puisque $p = q = \frac{1}{2}$, on peut encore écrire : $\mathrm{P}\{T \leq 2k\} = \mathrm{P}\{Y_{2k} > 1\} + \mathrm{P}\{Y_{2k} < -1\}\mathrm{P}\{Y_{2k} = 1\} = 1 - \mathrm{P}\{Y_{2k} = -1\}$. Cette dernière probabilité s'évalue par un simple comptage : $\binom{2k+1}{k+1}2^{-2k}$. Par conséquent, $\mathrm{P}\{T \leq 2k\} = 1 - \binom{2k+1}{k+1}2^{-2k}$.

2.6 **(a)** On suppose les vignettes numérotées de 1 à m; pour chaque $n \geq 1$, on désigne par Y_n le numéro de la vignette obtenue au $n^{\text{ième}}$ achat. Les Y_n sont supposées mutuellement indépendantes et chacune uniformément répartie sur $\{1, 2, \ldots, m\}$. Les variables aléatoires U_1, U_2, ... , U_m sont à valeurs dans $\{1, 2, \ldots\}$. Pour toute suite $(u_1 = 1, u_2, \ldots, u_m)$ d'entiers, l'évènement $\{U_1 = u_1, U_2 = u_2, \ldots, U_m = u_m\}$ est réalisé si et seulement s'il existe une *permutation* $(j_1, j_2, \ldots, j_m)$ de $(1, 2, \ldots, m)$ (les numéros des vignettes quand elles apparaissent pour la première fois) telle que soit réalisée la conjonction des évènements : $\{Y_1 = j_1, Y_2 = j_1, \ldots, Y_{u_2} = j_1, Y_{u_2+1} = j_2\}$,
$\{Y_{u_2+2}, \ldots, Y_{u_2+u_3} \in \{j_1, j_2\}, Y_{u_2+u_3+1} = j_3\}$,
$\{Y_{u_2+u_3+2}, \ldots, Y_{u_2+u_3+u_4} \in \{j_1, j_2, j_3\}, Y_{u_2+u_3+u_4+1} = j_4\}$, ...
$\{Y_{u_2+\cdots+u_{m-1}+2}, \ldots, Y_{u_2+\cdots+u_m} \in \{j_1, \ldots, j_{m-1}\}, Y_{u_2+\cdots+u_m+1} = j_m\}$. On en tire :

$$\begin{aligned}\mathrm{P}\{U_1 = u_1, U_2 = u_2, \ldots, U_m = u_m\}&\\ &= m!\Big(\frac{1}{m}\Big)^{u_2}\Big(\frac{1}{m}\Big)\cdot\Big(\frac{2}{m}\Big)^{u_3-1}\Big(\frac{1}{m}\Big)\cdots\Big(\frac{m-1}{m}\Big)^{u_m-1}\Big(\frac{1}{m}\Big)\\ &= \Big(\frac{1}{m}\Big)^{u_2-1}\Big(\frac{m-1}{m}\Big)\cdot\Big(\frac{2}{m}\Big)^{u_3-1}\Big(\frac{m-2}{m}\Big)\cdots\Big(\frac{m-1}{m}\Big)^{u_m-1}\Big(\frac{1}{m}\Big).\end{aligned}$$

(b) Si on somme l'identité ci-dessus par rapport à tous les u_i ($i = 2, \dots, m$) sauf l'un d'eux, disons k ($2 \le k \le m$), on trouve : $\mathrm{P}\{U_k = u_k\} = \left(\frac{k-1}{m}\right)^{u_k-1}\left(\frac{m-k+1}{m}\right)$ ($u_k = 1, 2, \dots$) ; ce qui montre que U_k suit une loi géométrique de paramètre $p = (m-k+1)/m$ et que les U_k sont mutuellement indépendantes.

(c) La fonction génératrice d'une variable géométrique de paramètre p est $ps(1-qs)^{-1}$. La fonction génératrice de U_k est donc :

$$G_{U_k}(s) = \frac{m-k+1}{m}\, s\left(1 - \frac{k-1}{m}s\right)^{-1} \quad (k = 1, 2, \dots, m).$$

Or $T = T_m = (T_m - T_{m-1}) + (T_{m-1} - T_{m-2}) + \dots + (T_2 - T_1) + T_1 = U_1 + U_2 + \dots + U_m$. La fonction génératrice de T est donc

$$G_T(s) = \frac{m!}{m^m}\, s^m \prod_{1\le k\le m}\left(1 - \frac{k-1}{m}s\right)^{-1}.$$

L'espérance d'une variable géométrique de paramètre p est $1/p$, d'où

$$\mathbb{E}[T] = \sum_{1\le k\le m} \frac{m}{m-k+1} = m\left(1 + \frac{1}{2} + \dots + \frac{1}{m}\right).$$

On retiendra que T est la somme de m variables aléatoires géométriques indépendantes, de paramètres respectifs m/m, $(m-1)/m$, ... , $1/m$. La variable T est donc presque sûrement *finie*.

Autre variante : on peut calculer directement la fonction génératrice de T à partir de la fonction génératrice des nombres de Stirling de seconde espèce. L'évènement $\{Y_1 = j_1, \dots, Y_n = j_n\}$ est contenu dans $\{T = n\}$ si et seulement si l'application $k \mapsto j_k$ ($k = 1, 2, \dots, n-1$) est une surjection de l'intervalle $[1, n-1]$ sur l'ensemble $[1, m] \setminus \{j_n\}$. D'où $\mathrm{P}\{T = n\} = (m-1)!\, S(n-1, m-1)\, m/m^n$, où $S(n,k)$ désigne le nombre de Stirling de seconde espèce.[1] Or, (*cf. op. cit.*) $\sum_{n\ge m} S(n-1, m-1)\, s^{n-m} = \prod_{1\le k\le m-1} (1-ks)^{-1}$. D'où l'on tire : $G_T(s) = \sum_{n\ge m} \mathrm{P}\{T = n\}\, s^n = m!\,(s/m)^m \sum_{n\ge m} S(n-1, m-1)(s/m)^{n-m}$ $= m!\,(s/m)^m \prod_{1\le k\le m} (1 - (k-1)s/m)^{-1}$.

2.7 **(a)** Conditionnellement à l'évènement $\{Y_1 = 0\}$ (resp. à l'évènement $\{Y = 1, Y_2 = 1\}$, resp. à l'évènement $\{Y_1 = 1, Y_2 = 0\}$), la loi de N est celle de $1 + N$ (resp. la loi de la constante 2, resp. la loi de $2 + N$). D'où $\mathbb{E}[N \mid Y_1 = 0] = 1 + \mathbb{E}[N]$, puis $\mathbb{E}[N \mid Y_1 = 1, Y_2 = 1] = 2$, enfin $\mathbb{E}[N \mid Y_1 = 1, Y_2 = 0] = 2 + \mathbb{E}[N]$. Comme $\{Y_1 = 0\}$, $\{Y = 1, Y_2 = 1\}$ et $\{Y_1 = 1, Y_2 = 0\}$ forment un système complet d'évènements, on a : $\mathbb{E}[N] = \mathbb{E}[N \mid Y_1 = 0]\,\mathrm{P}\{Y_1 = 0\} + \mathbb{E}[N \mid Y_1 = 1, Y_2 = 1]$

[1] Comtet (Louis). — *Analyse Combinatoire*, vol. 2, p. 42. — Paris, Presses Universitaires de France, 1970.

$\mathrm{P}\{Y_1 = 1, Y_2 = 1\} + \mathbb{E}[N \mid Y_1 = 1, Y_2 = 0]\,\mathrm{P}\{Y_1 = 1, Y_2 = 0\} = (1 + \mathbb{E}[N])\frac{1}{2} + 2\,\frac{1}{4} + (2 + \mathbb{E}[N])\frac{1}{4}$, soit $\mathbb{E}[N] = 6$ enfants et donc 3 garçons en moyenne.

(b) Conditionnellement à l'évènement $\{Y_1 = 1\}$, la loi de N est celle de $1 + N$. Conditionnellement à $\{Y_1 = 0\}$, la loi de N est la loi de $1 + Z$, où Z est une variable aléatoire géométrique de paramètre $\frac{1}{2}$: $\mathrm{P}\{Z = k\} = (1/2)(1/2)^{k-1} = (1/2)^k$ pour $k = 1, 2, \ldots$ D'où $\mathbb{E}[Z] = 2$ et $\mathbb{E}[N] = \mathbb{E}[N \mid Y_1 = 1]\,\mathrm{P}\{Y_1 = 1\} + \mathbb{E}[N \mid Y_1 = 0]\,\mathrm{P}\{Y_1 = 0\} = (1 + \mathbb{E}[N])\frac{1}{2} + (1 + \mathbb{E}[Z])\,\frac{1}{2}$, soit $\mathbb{E}[N] = 4$ enfants et donc 2 garçons en moyenne.

On peut également évaluer la fonction génératrice $G(s)$ de N dans les deux exemples. Posons $u_n := \mathrm{P}\{N = n\}$. Dans le cas (a), $u_1 = 0$, $u_2 = \frac{1}{4}$ et $u_n = \frac{1}{2}u_{n-1} + \frac{1}{4}u_{n-2}$ pour $n \geq 3$. On en tire : $G(s) = \frac{1}{4}s^2 + \frac{1}{2}s\,G(s) + \frac{1}{4}s^4\,G(s)$, d'où $G(s) = \frac{1}{4}s^2(1 - \frac{1}{2}s - \frac{1}{4}s^2)^{-1}$ et $\mathbb{E}[N] = G'(1) = 6$. Dans le cas (b), $u_1 = 0$ et $u_n = \frac{1}{2}u_{n-1} + (\frac{1}{2})^n$ pour $n \geq 2$. D'où $G(s) = \frac{1}{2}s\,G(s) + (\frac{1}{2}s)^2(1 - \frac{1}{2}s)^{-1}$ et, par conséquent, $G(s) = (\frac{1}{2}s)^2(1 - \frac{1}{2}s)^{-2}$, puis $\mathbb{E}[N] = G'(1) = 4$.

2.8 **(a)** Le processus (X_n) $(n \geq 0)$ prend ses valeurs dans $E = \{-1, 0, +1\}$ et T est le *premier* instant où il entre dans $A := \{-1, +1\}$. C'est un temps d'atteinte, donc un temps d'arrêt adapté au processus (X_n) $(n \geq 0)$. Pour chaque $n \geq 0$, on a $\mathrm{P}\{X_n \in A\} = 2pq$. Pour tout $k \geq 0$, on en déduit $\mathrm{P}\{T = k\} = \{X_0 = 0, \ldots, X_{k-1} = 0, X_k \in A\} = (1 - 2pq)^k\,2pq$. Donc T suit la loi géométrique de paramètre $2pq$ et $\mathbb{E}[T] = 1/(2pq)$.

(b) La variable X_T prend ses valeurs dans A. Puisque T est presque sûrement finie, $\mathrm{P}\{X_T = 1\} = \sum_{k \geq 1} \mathrm{P}\{X_T = 1 \mid T = k\}\,\mathrm{P}\{T = k\} = \sum_{k \geq 1} \mathrm{P}\{X_k = 1 \mid T = k\}\,\mathrm{P}\{T = k\}$. Pour tout $k \geq 1$, on a : $\mathrm{P}\{X_k = 1 \mid T = k\} = \mathrm{P}\{X_k = 1 \mid X_1 = 0, \ldots, X_{k-1} = 0, X_k \in A\}$. Comme $X_1, \ldots, X_k$ sont indépendantes, cette dernière probabilité vaut : $\mathrm{P}\{X_k = 1 \mid X_k \in A\} = \mathrm{P}\{X_k = 1, X_k \in A\}/\mathrm{P}\{X_k \in A\} = \mathrm{P}\{X_k = 1\}/\mathrm{P}\{X_k \in A\} = pq/(2pq) = \frac{1}{2}$. D'où $\mathrm{P}\{X_T = 1\} = \frac{1}{2}\sum_{k \geq 1} \mathrm{P}\{T = k\} = \frac{1}{2}\mathrm{P}\{T < +\infty\} = \frac{1}{2}$. De même, on montre que $\mathrm{P}\{T = -1\} = \frac{1}{2}$. La loi de X_T est donc $\frac{1}{2}(\varepsilon_{-1} + \varepsilon_{+1})$.

CHAPITRE 3

3.1 Pour $0 \leq s \leq t$, on a : $\mathrm{P}\{T_1 \in [0, s] \mid N(t) = 1\} = \mathrm{P}\{T_1 \in [0, s], N(t) = 1\}/\mathrm{P}\{N(t) = 1\} = \mathrm{P}\{\text{un top dans } [0, s],\ 0 \text{ top dans }]s, t]\}/\mathrm{P}\{N(t) = 1\} = \mathrm{P}\{\text{un top dans } [0, s]\}\mathrm{P}\{0 \text{ top dans }]s, t]\}/\mathrm{P}\{N(t) = 1\} = e^{-\lambda s}\lambda s e^{-\lambda(t-s)}/(e^{-\lambda t}\lambda t) = s/t$.

3.2 Le premier greffon arrive après un temps de T_1 unités, le second après $T_1 + T_2$. Soit X_1 (resp. X_2) la durée de vie du premier (resp. du second patient). Les variables aléatoires T_1, T_2, X_1, X_2 sont indépendantes, exponentielles, les deux premières de paramètre λ, les deux dernières de

paramètre μ_1 et μ_2, respectivement. La probabilité que le premier patient soit greffé est égale à $\mathrm{P}\{X_1 > T_1\} = \lambda/(\lambda+\mu_1)$, d'après un calcul qui a été fait dans [FF1], Exemple 2, chap. 12, § 4. On peut aussi évaluer l'intégrable double $\iint_{0<t<x} \mu_1 e^{-\mu_1 x}\, \lambda e^{-\lambda t}\, dx\, dt$.

Le second patient est greffé si la réunion des deux évènements suivants est réalisée : $\{X_1 < T_1 < X_2\} + \{T_1 < X_1, T_1 + T_2 < X_2\}$. Or, $\mathrm{P}\{X_1 < T_1 < X_2\} = \iiint_{0<x<t<y} \mu_1 e^{-\mu_1 x}\, \lambda e^{-\lambda t}\, \mu_2 e^{-\mu_2 y}\, dx\, dt\, dy$, une intégrale qui vaut $\lambda\mu_1/(\lambda+\mu_2)(\lambda+\mu_1+\mu_2)$. Puis, $\mathrm{P}\{T_1 < X_1, T_1+T_2 < X_2\} = \iiiint_{0<t_1<x, t_1+t_2<y} \lambda e^{-\lambda t_1} \mu_1 e^{-\mu_1 x}\, \lambda e^{-\lambda t_2} \mu_2 e^{-\mu_2 y}\, dt_1\, dx\, dt_2\, dy$, une intégrale qui vaut $(\lambda/(\lambda+\mu_1+\mu_2)) \cdot (\lambda/(\lambda+\mu_2))$. (Intégrer d'abord par rapport à x et à y.)

3.3 **(a)** En effet, $X(t) = \sum_{i=1}^{N(t)} (t - S_i) = t\,N(t) - \sum_{i=1}^{N(t)} S_i$.

(b) Conditionnellement à $\{N(t) = n\}$ le vecteur $(S_1, \ldots, S_n)$ a même loi que l'échantillon ordonné $(U_1^*, \ldots, U_n^*)$ de n variables aléatoires $U_1, \ldots, U_n$, indépendantes, identiquement distribuées, de loi uniforme sur $[0, t]$. Donc $\mathbb{E}[\sum_{i=1}^{N(t)} S_i \mid N(t) = n] = \mathbb{E}[\sum_{i=1}^{n} S_i \mid N(t) = n] = \mathbb{E}[\sum_{i=1}^{n} U_i^*] = \mathbb{E}[\sum_{i=1}^{n} U_i] = nt/2$.
D'où $\mathbb{E}[\sum_{i=1}^{n} S_i] = \sum_{n\ge 0} \mathbb{E}[\sum_{i=1}^{n} S_i \mid N(t) = n] \times \mathrm{P}\{N(t) = n\} =$
$(t/2) \sum_{n\ge 0} n\,\mathrm{P}\{N(t) = n\} = (t/2)\,\mathbb{E}[N(t)] = \lambda\,t^2/2$.
Soit, enfin, $\mathbb{E}[X(t)] = t\,\mathbb{E}[N(t)] - \mathbb{E}[\sum_{i=1}^{n} S_i] = \lambda\,t^2/2$.

3.4 On peut écrire $\mathbb{E}[A(t) \mid N(t) = n] = \mathbb{E}[\sum_{k=1}^{n} A_k\, e^{-\alpha(t-X_k)} \mid N(t) = n] = \sum_{k=1}^{n} \mathbb{E}[A_k]\,\mathbb{E}[e^{-\alpha(t-X_k)} \mid N(t) = n] = n\,\mathbb{E}[A]\,\mathbb{E}[e^{-\alpha(t-X)}]$, en faisant appel à la Proposition 3.5.3, la variable X étant uniformément répartie sur $[0, t]$.
Or, $\mathbb{E}[e^{-\alpha(t-X)}] = (1/t) \int_0^t e^{-\alpha(t-x)}\, dx = (1 - e^{-\alpha t})/\alpha\, t$. D'où
$\mathbb{E}[A(t) \mid N(t) = n] = n\,\mathbb{E}[A]\,(1 - e^{-\alpha t})/\alpha\, t$ et donc
$\mathbb{E}[A(t) \mid N(t)] = N(t)\,\mathbb{E}[A]\,(1 - e^{-\alpha t})/\alpha\, t$ et $\mathbb{E}[A(t)] = \mathbb{E}[\mathbb{E}[A(t) \mid N(t)]] =$
$\mathbb{E}[N(t)]\,\mathbb{E}[A](1 - e^{-\alpha t})/\alpha\, t = \lambda\,\mathbb{E}[A]\,(1 - e^{-\alpha t})/\alpha$.

3.5 Supposons $0 \le s < t$. Alors $\mathbb{E}[N(s)\,N(t)] = \mathbb{E}[N(s)(N(s) + (N(t) - N(s)))] = \mathbb{E}[N(s)^2] + \mathbb{E}[N(s)(N(t) - N(s))]$, d'où, puisque le processus est à accroissements indépendants,
$\mathbb{E}[N(s)\,N(t)] = \mathbb{E}[N(s)^2] + \mathbb{E}[N(s)]\,\mathbb{E}[(N(t) - N(s))]$. On en tire
$\mathrm{Cov}(N(s), N(t)) = \mathbb{E}[N(s)\,N(t)] - \mathbb{E}[N(s)]\,\mathbb{E}[N(t)]$
$= \mathbb{E}[N(s)^2] + \mathbb{E}[N(s)](\mathbb{E}[(N(t) - N(s))] - \mathbb{E}[N(t)]) = \mathbb{E}[N(s)^2] - (\mathbb{E}[N(s)])^2 =$
$\mathrm{Var}\,N(s) = \lambda\, s$. En général, $\mathrm{Cov}(N(s), N(t)) = \lambda\, \min(s, t)$.

3.6 Pour tout $t \ge 0$ et tout $h > 0$, on a : $\mathbb{E}[(N(t+h) - N(t))^2] = \mathbb{E}[N(h)^2] = \mathrm{Var}\,N(h) + \mathbb{E}[N(h)]^2 = \lambda\, h + (\lambda\, h)^2 = \lambda h + o(h)$, une quantité qui tend vers 0 lorsque h tend vers 0.

3.7 Pour $\varepsilon > 0$, $h > 0$ tels que $\varepsilon h \le 1$, on a : $\mathrm{P}\Big\{\dfrac{N(t+h)-N(t)}{h} \ge \varepsilon\Big\} = \mathrm{P}\{N(t+h)-N(t) \ge \varepsilon h\} = \mathrm{P}\{N(h) \ge \varepsilon h\} \le \mathrm{P}\{N(h) \ge 1\} = \lambda h + o(h)$, qui tend vers 0, lorsque h tend vers 0. En revanche, $\mathbb{E}\Big[\Big(\dfrac{N(t+h)-N(t)}{h}\Big)^2\Big] = \dfrac{1}{h^2}\,\mathbb{E}[N(h)^2] = \dfrac{1}{h^2}(\lambda h + (\lambda h)^2)$, une quantité qui tend vers l'infini lorsque h tend vers 0.

3.8 Conditionnellement à l'évènement $\{T = t\}$ $(t > 0)$, la variable aléatoire Z suit la loi de Poisson, de paramètre λt. Pour tout $n \ge 0$, on obtient donc : $\mathrm{P}\{Z = n\} = \int_0^{+\infty} \mathrm{P}\{Z = n \mid T = t\}\, f(t)\, dt$, où $f(t)$ désigne la densité de T, soit $\mu\, e^{-\mu t} I_{[0,+\infty[}$. D'où $\mathrm{P}\{Z = n\} = \int_0^{+\infty} e^{-\lambda t}((\lambda t)^n/n!)\, \mu\, e^{-\mu t}\, dt = (\lambda^n \mu/n!) \int_0^{+\infty} e^{-(\lambda+\mu)t}\, t^n\, dt$, d'où, par le changement de variables $(\lambda+\mu)t := u$, l'intégration : $\mathrm{P}\{Z = n\} = (\lambda^n\mu/n!) \int_0^{+\infty} e^{-u}\, (u/(\lambda+\mu))^n\, (1/(\lambda+\mu))\, du = (\lambda/(\lambda+\mu))^n\, (\mu/(\lambda+\mu)$. La variable aléatoire Z suit donc la loi géométrique, à valeurs dans $\{0, 1, \dots\}$, de paramètre $p = \mu/(\lambda+\mu)$. En particulier, $\mathrm{P}\{Z \ge 1\} = 1 - \mu/(\lambda+\mu) = \lambda/(\lambda+\mu)$.

3.9 En vertu de l'indépendance de X_0 et du processus, on a : $\mathbb{E}[X(t)] = \mathbb{E}[X_0]\,\mathbb{E}[(-1)^{N(t)}] = 0$. Ensuite $\mathrm{Cov}(X(s), X(t)) = \mathbb{E}[X(s)X(t)] =$
$\mathbb{E}[X_0^2\,(-1)^{N(s)+N(t)}] = \mathbb{E}[(-1)^{2N(s)+(N(t)-N(s))}] = \mathbb{E}[(-1)^{N(t)-N(s)}]$
$= \mathbb{E}[(-1)^{N(t-s)}]$, car le processus est à accroissements stationnaires. D'où $\mathrm{Cov}(X(s), X(t)) = \sum_{k\ge 0} (-1)^k\, e^{-\lambda(t-s)}\, \dfrac{\lambda(t-s))^k}{k!} = e^{-\lambda(t-s)} \sum_{k \ge 0} \dfrac{(-\lambda(t-s))^k}{k!} = e^{-2\lambda(t-s)}$.

3.10 **(a)** Dans tout cet exercice, on est ramené à calculer des densités marginales à partir de la densité conjointe de l'échantillon ordonné. D'abord,

$$\begin{aligned} f_{X_s^*}(x_s) &= \frac{n!}{t^n} \int_0^{x_s} dx_1 \cdots \int_{x_{s-2}}^{x_s} dx_{s-1} \times \int_{x_s}^{t} dx_{s+1} \cdots \int_{x_{n-1}}^{t} dx_n \\ &= \frac{n!}{t^n} \times \frac{x_s^{s-1}}{(s-1)!} \times \frac{(t-x_s)^{n-s}}{(n-s)!} \quad (0 < x_s < t). \end{aligned}$$

Pour trouver la densité de $t - X_s^*$, on effectue le changement de variable $y_s = t - x_s$.

(b) De même, on a :

$$\begin{aligned} f_{X_r^*, X_s^*}(x_r, x_s) = \frac{n!}{t^n} &\int_0^{x_r} dx_1 \cdots \int_{x_{r-2}}^{x_r} dx_{r-1} \times \int_{x_r}^{x_s} dx_{r+1} \\ &\cdots \int_{x_{s-2}}^{x_s} dx_{s-1} \times \int_{x_s}^{t} dx_{s+1} \cdots \int_{x_{n-1}}^{t} dx_n \\ = \frac{n!}{t^n} &\times \frac{x_r^{r-1}}{(r-1)!} \times \frac{(x_s - x_r)^{s-r-1}}{(s-r-1)!} \times \frac{(t-x_s)^{n-s}}{(n-s)!} \\ &(0 < x_r < x_s < t). \end{aligned}$$

(c) Lorsque $r = 0$ ou $s = n + 1$, la densité de $U_{r,s}$ a été calculée en (a). Avec $1 \le r < s \le n$, on part de la densité $f_{X^*_r, X^*_s}(x_r, x_s)$, calculée en (b) et on effectue le changement de variables : $u := x_s - x_r$, $v := x_r$. On obtient

$$\binom{n}{r-1, 1, s-r-1, 1, n-s} \frac{1}{t^n} v^{r-1} u^{s-r-1} (t-u-v)^{n-s},$$

que l'on intègre de $v = 0$ à $v = t - u$. Avec $v := (t-u)y$, on a :

$$\begin{aligned}\int_0^{t-u} v^{r-1}(t-u-v)^{n-s}\,dv &= (t-u)^{n-s+r}\int_0^1 y^{r-1}(1-y)^{n-s}\,dy \\ &= (t-u)^{n-s+r} B(r, n-s+1) \\ &= (t-u)^{n-s+r}\frac{(r-1)!\,(n-s)!}{(n-s+r)!}\end{aligned}$$

(*cf.* [FF1], chap. 14, § 9). En reportant, on trouve la densité de $U_{r,s}$. Noter que l'expression trouvée se spécialise en les formules obtenues en (a), lorsque $r = 0$ ou $s = n + 1$.

(d) C'est la spécialisation de (c) pour $r = 1$ et $s = n$.

(e) C'est la spécialisation de (c) pour $s = r + 1$. Il est remarquable que cette densité ne dépend pas de r.

3.11 **(a)** Par définition-même, $Q(t) = \sum_{k=2}^{N(t)+1} I_{\{T_k > l\}}$ est une variable qui a même loi que $\sum_{k=1}^{N(t)} I_{\{T_k > l\}}$, puisque les T_i sont identiquement distribuées.

(b) On a : $\mathbb{E}[Q(t)] = \sum_{n\ge1} \mathbb{E}[\sum_{k=1}^{N(t)} I_{\{T_k>l\}} \mid N(t) = n]\,\mathrm{P}\{N(t) = n\} = \sum_{n\ge1} \mathbb{E}[\sum_{k=1}^{n} I_{\{T_k>l\}} \mid N(t) = n]\,\mathrm{P}\{N(t) = n\} = \sum_{n\ge1} \mathbb{E}[\sum_{k=1}^{n} I_{\{U_{k-1,k}>l\}}]\,\mathrm{P}\{N(t) = n\} = \sum_{n\ge1}\sum_{k=1}^{n} \mathrm{P}\{U_{k-1,k} > l\}\,\mathrm{P}\{N(t) = n\} = \sum_{n\ge1} n(1/t^n)(t-l)^n e^{-\lambda t}(\lambda t)^n/n! = \lambda(t-l)e^{-\lambda l}\sum_{n\ge1}(t-l)^{n-1}e^{-\lambda(t-l)}/(n-1)! = \lambda(t-l)e^{-\lambda l}$.

(c) Cette formule n'a d'intérêt que si $0 \le l \le t$. On obtient :
$\mathbb{E}[Q(t)] = \begin{cases}\lambda t, & \text{si } l = 0,\\ 0, & \text{si } l = t.\end{cases}$

CHAPITRE 4

4.1 **(a)** Soit $N(t)$ le nombre de véhicules passant dans l'intervalle $[0, t]$. On suppose que $\{N(t) : t \ge 0\}$ est un processus de Poisson de densité λ, qui est l'espérance mathématique du nombre de passages par unité de temps, disons la minute, soit $\lambda = 6$. Le nombre de camions passant au point fixé dans un intervalle de temps est un processus de Poisson de densité

$\lambda p = 6 \times 0,1 = 0,6$. La probabilité qu'au moins un camion passe en une minute est : $1 - e^{-0,6} \approx 0,45$.

(b) Soit $N_c(t)$ (resp. $N_v(t)$) le processus de Poisson régissant le flot des camions (resp. des voitures). Il s'agit de calculer $\mathbb{E}[N(5) \,|\, N_c(5) = 10]$, qui vaut $\mathbb{E}[N_c(5) \,|\, N_c(5) = 10] + \mathbb{E}[N_v(5) \,|\, N_c(5) = 10] = 10 + \mathbb{E}[N_v(5)] = 10 + 6 \times 0,9 \times 5 = 37$.

(c) Conditionnellement à $\{N(10) = 30\}$ les arrivées des trente véhicules se répartissent uniformément sur $[0, 10]$, avec une proportion de camions égale à $p = 1/10$. Le nombre de camions suit alors une loi binomiale de paramètre $n = 30, p = 1/10$. La probabilité cherchée est de : $\binom{30}{3}(1/10)^3(9/10)^{27} \approx 0,24$.

4.2 On suppose que les instants d'arrivée des citoyens (resp. des étrangers) sont régis par un processus de Poisson, de paramètre λ (resp. μ). On suppose ces processus indépendants. Soit $S_n^{(c)}$ (resp. $S_m^{(e)}$) l'instant d'arrivée du $n^{\text{ième}}$ citoyen (resp. du $m^{\text{ième}}$ étranger). Il s'agit de calculer $\mathrm{P}\{S_n^{(c)} < S_m^{(e)}\}$. Comme $S_n^{(c)}$ (resp. $S_m^{(e)}$) suit une loi gamma de paramètre (n, λ) (resp. (m, μ)) et que ces variables sont indépendantes, on a :

$$\begin{aligned}
\mathrm{P}\{S_n^{(c)} < S_m^{(e)}\} &= \int_0^{+\infty} \frac{\lambda}{(n-1)!}\, e^{-\lambda x}(\lambda x)^{n-1}\, dx \int_x^{+\infty} \frac{\mu}{(m-1)!} e^{-\mu y}(\mu y)^{m-1}\, dy \\
&= \int_0^{+\infty} \frac{\lambda}{(n-1)!}\, e^{-\lambda x}(\lambda x)^{n-1} \sum_{k=0}^{m-1} e^{-\mu x}\frac{(\mu x)^k}{k!}\, dx \\
&= \sum_{k=0}^{m-1} \binom{n+k-1}{k} p^n\, q^k,
\end{aligned}$$

avec $p := \lambda/(\lambda + \mu)$, $q = 1 - p$. On voit alors que $\mathrm{P}\{S_n^{(c)} < S_m^{(e)}\}$ est la probabilité que, dans une suite indépendante de jets de « pile » ou « face », avec p comme probabilité d'obtenir « pile », on obtienne n « pile » avant m « face », en au plus $(n + m - 1)$ jets. On peut évaluer cette probabilité par un autre comptage et on en déduit l'identité :

$$\sum_{k=0}^{m-1} \binom{n+k-1}{k} p^n\, q^k = \sum_{k=n}^{n+m-1} \binom{n+m-1}{k} p^k q^{n+m-1-k}.$$

On peut également arriver à la toute dernière formule, en utilisant l'argument de la seconde Remarque de la Proposition 4.2.1. On considère la somme $\{N(t) : t \geq 0\}$ des processus $\{N^{(c)}(t) : t \geq 0\}$ et $\{N^{(e)}(t) : t \geq 0\}$. La probabilité cherchée est la probabilité que lors de l'apparition des $(n + m - 1)$ premiers tops, il y en ait au moins n qui proviennent du processus $\{N^{(c)}(t) : t \geq 0\}$.

L'identité précédente obtient ci-dessus une démonstration probabiliste. Elle peut, toutefois, se démontrer directement de la façon suivante. En utilisant la notation des factorielles montantes, le membre de gauche se récrit

$p^n \sum_{k=0}^{m-1}((n)_k/k!)q^k$ et le membre de droite $p^n \sum_{l=0}^{m-1}\binom{n+m-1}{l}p^{m-1-l}q^l =$ $p^n \sum_{l=0}^{m-1}(-1)^l \frac{(-n-m+1)_l}{l!} q^l(1-q)^{n-1-l}$. Le coefficient de q^k dans cette dernière somme vaut $\sum_{l=0}^{k}(-1)^l \frac{(-n-m+1)_l}{l!}\frac{(-m+1+l)_{k-l}}{(k-l)!}$. Comme $\binom{k}{l} =$ $(-1)^l \frac{(-k)_l}{l!}$ et $(-m+1)_k = (-m+1)_l\,(-m+1+l)_{k-l}$, ce coefficient vaut encore $\frac{(-m+1)_k}{k!}\sum_{l=0}^{k}\frac{(-k)_l\,(-n-m+1)_l}{l!\,(-m+1)_l} = \frac{(n)_k}{k!}$, en utilisant l'identité de Chu-Vandermonde (*cf.* chap. 1, § 2). Le membre de droite est donc aussi égal à $p^n \sum_{k=0}^{m-1}((n)_k/k!)q^k$.

4.3 Comme dans la démonstration du Théorème 4.4.1, on a : $\mathrm{P}\{S(t) \le x\} = \sum_{n\ge 0} \mathrm{P}\{\sum_{k=1}^{N(t)} X_k \le x, N(t)=n\} = \sum_{n\ge 0} \mathrm{P}\{\sum_{k=1}^{n} X_k \le x, N(t)=n\} =$ $\sum_{n\ge 0} \mathrm{P}\{\sum_{k=1}^{n} X_k \le x\}\,\mathrm{P}\{N(t)=n\} = \sum_{n\ge 0} \mathrm{F}^{*n}(x)\,e^{-\lambda t}\,(\lambda\, t)^n/n!$

4.4 **(a)** En effet, l'évènement $\{Y_{n+2m} = m\}$ est réalisé si et seulement si, parmi $X_1, X_2, \ldots, X_{2n+m}$, il y a exactement $(n+m)$ variables qui valent $+1$ et n qui valent -1. Il y a aussi $\binom{2n+m}{n}$ d'obtenir cette situation.

(b) En utilisant le système complet $\{N(t)=n\}$ $(n \ge 0)$, on obtient :

$$\mathrm{P}\{S(t)=m\} = \sum_{n=0}^{\infty} \mathrm{P}\{\sum_{k=1}^{N(t)} X_k = m, N(t)=n\} =$$

$$\sum_{n=0}^{\infty} \mathrm{P}\{\sum_{k=1}^{n} X_k = m, N(t)=n\} = \sum_{n=0}^{\infty} \mathrm{P}\{Y_n = m\}\mathrm{P}\{N(t)=n\} =$$

$$\sum_{n=0}^{\infty} \mathrm{P}\{Y_{2n+m} = m\}\mathrm{P}\{N(t)=2n+m\} = \sum_{n=0}^{\infty}\binom{2n+m}{n}p^{n+m}q^n\frac{e^{-\lambda t}(\lambda t)^{2n+m}}{(2n+m)!}.$$

(c) On a : $\sum_{n,m} \mathrm{P}\{S(t)=m, N(t)=n\} = e^{-\lambda t}\sum_{n\ge 0}\frac{1}{n!}(pq\lambda^2t^2)^n$

$$\times \sum_{m\ge -n}\frac{p^m(\lambda t)^m}{(n+m)!} = e^{-\lambda t}\sum_{n\ge 0}\frac{1}{n!}(pq\lambda^2t^2)^n\sum_{k\ge 0}\frac{p^{k-n}(\lambda t)^{k-n}}{k!} = e^{-\lambda t}e^{-\lambda qt}e^{-\lambda pt} = 1.$$

4.5 **(a)** Pour $k \ge 0$, on a : $\{N_1(t) = k\} = \{N(t) = 2k\} \cup \{N(t) = 2k+1\}$. D'où $\mathrm{P}\{N_1(t)=k\} = e^{-\lambda t}(\lambda t)^{2k}/(2k)! + e^{-\lambda t}(\lambda t)^{2k+1}/(2k+1)! =$ $(e^{-\lambda t}(\lambda t)^{2k}/(2k)!)(1+\lambda t/(2k+1))$.

(b) Si $t < a$, on a $\mathrm{P}\{N_2(t) = 0\} = 1$ et $\mathrm{P}\{N_2(t) = k\} = 0$ pour $k \geq 1$. Pour $t \geq a$ et $k \geq 0$ il vient : $\mathrm{P}\{N_2(t) = k\} = \mathrm{P}\{N(t-a) = k\} = e^{-\lambda t}e^{\lambda a}((\lambda(t-a))^k/k!$

(c) Soit X l'instant où s'est produit l'un des n tops et soit Y la quantité aléatoire, de densité $g(x)$, que l'on ajoute à cet X. Conditionnellement à $\{N(t) = n\}$, la variable $N_3(t)$ a même loi que la somme $I_{\{X_1+Y_1\leq t\}} + \cdots + I_{\{X_n+Y_n\leq t\}}$, qui suit une loi binomiale de paramètres (n, p). Il s'agit d'évaluer p. Or, $p = \mathbb{E}[I_{\{X+Y\leq t\}}] = \mathrm{P}\{X+Y \leq t\} = \int_{\mathbb{R}+} \mathbb{E}[I_{\{X+Y\leq t\}}\,|\,X = x](1/t)I_{[0,t]}(x)\,dx$ et $\mathbb{E}[I_{\{X+Y\leq t\}}\,|\,X = x] = \int_{\mathbb{R}+} I_{\{x+y\leq t\}}(y)\,g(y)\,dy$, qui vaut $G(t-x) = \int_0^{t-x} g(y)\,dy$ si $t > x$ et 0 si $t \leq x$. D'où $p = \int_{\mathbb{R}+} G(t-x)(1/t)I_{[0,t]}(x)\,dx = (1/t)\int_0^t G(t-x)\,dx$. On en déduit que $\mathrm{P}\{N_3(t) = k\,|\,N(t) = n\}$ est nul si $n < k$ et vaut $\binom{n}{k}p^k(1-p)^{n-k}$ si $n \geq k$. Par conséquent, $\mathrm{P}\{N_3(t) = k\} = \sum_{n\geq k}\binom{n}{k}p^k(1-p)^{n-k}\frac{e^{-\lambda t}(\lambda t)^n}{n!} = \frac{(\lambda tp)^k e^{-\lambda tp}}{k!}$. Noter que, dans (b), on a pris $G(t) = 1$ si $t \geq a$ et $G(t) = 0$ si $t < a$.

(d) Ici $V_n = (T_{n-1} + T_n)/2$, d'où $\mathbb{E}[V_nV_{n+1}] = (1/4)\mathbb{E}[T_{n-1}T_n + T_{n-1}T_{n+1} + T_n^2 + T_nT_{n+1}] = (1/4)(\mathbb{E}[T_1^2] + 3\,\mathbb{E}[T_1]^2) = (1/4)((2/\lambda^2) + 3(1/\lambda)^2) = 5/(4\lambda^2)$ et $\mathbb{E}[V_n] = \mathbb{E}[V_{n+1}] = 1/\lambda$. Par ailleurs, $\mathrm{Var}\,V_n = (1/4)(\mathrm{Var}\,T_{n-1} + \mathrm{Var}\,T_n) = 1/(2\lambda^2)$. L'écart-type vaut $\sigma_{V_n} = 1/(\lambda\sqrt{2})$ et le coefficient de corrélation linéaire $\rho(V_n, V_{n+1}) = (5/(4\lambda^2) - 1/\lambda^2)/(1/(2\lambda^2)) = 1/2$. De plus, $\rho(V_n, V_{n+2}) = 0$. Enfin, $\mathrm{P}\{V_n \leq x\} = \mathrm{P}\{T_{n-1} + T_n \leq 2x\} = \int_0^{2x}(\lambda/\Gamma(2))e^{-\lambda u}(\lambda u)\,du$.

4.6 **(a)** En effet, $P_n(t+h) = \mathrm{P}\{N(t+h) = n\} = \mathrm{P}\{N(t) = n\}\mathrm{P}\{N(t+h) - N(t) = 0\} + \mathrm{P}\{N(t) = n-1\}\mathrm{P}\{N(t+h) - N(t) = 1\} + \sum_{2\leq k\leq n}\mathrm{P}\{N(t) = n-k\}\mathrm{P}\{N(t+h) - N(t) = k\} = P_n(t)(1 - \lambda h + o(h)) + P_{n-1}(t)(\lambda h + o(h)) + o(h) = P_n(t)(1 - \lambda h) + P_{n-1}(t)(\lambda h + o(h))$. D'où $\frac{P_n(t+h) - P_n(t)}{h} = -\lambda\,P_n(t) + \lambda\,P_{n-1}(t) + \frac{o(h)}{h}$. En faisant tendre h vers 0, on trouve $P_n'(t) = -\lambda\,P_n(t) + \lambda P_{n-1}(t)$ pour $n = 1, 2, \ldots$; de plus, $P_n(0) = 0$ pour $n = 1, 2, \ldots$

(b) Avec la même décomposition que dans (a), on obtient : $P_n(t+h) = P_n(t)(1 - \lambda_n o(h) + o(h)) + P_{n-1}(t)(\lambda_{n-1}h + o(h))$. En faisant tendre h vers 0, on trouve : $P_n'(t) = -\lambda_n\,P_n(t) + \lambda_{n-1}\,P_{n-1}(t)$ pour $n = 1, 2, \ldots$ Pour $n = 0$, il vient $P_0(t+h) = \mathrm{P}\{N(t+h) = 0\} = \mathrm{P}\{N(t) = 0\} \times \mathrm{P}\{N(t+h) - N(t) = 0\} = P_0(t)(1 - \lambda_0 h + o(h))$. D'où $P_0'(t) = -\lambda_0\,P_0(t)$.

(c) Si $n_0 = 1$, alors $P_0'(t) = 0$ et $P_0(t)$ est constant, mais comme $P_0(0) = \mathrm{P}\{N(0) = 0\} = 0$, cette constante est nulle. Pour $n \geq 0$, posons $Q_n(t) := P_n(t)e^{n\lambda t}$. Le système différentiel, satisfait par les $P_n(t)$, devient pour les $Q_n(t)$: $Q_n'(t) = (n-1)\lambda e^{\lambda t}Q_{n-1}(t)$ pour $n \geq 1$ et $Q_0'(t) = -\lambda_0 Q_0(t)$. De plus, $Q_n(0) = 1$ si $n = 1$ et $Q_n(0) = 0$ si $n \geq 2$. Par intégration, on obtient, par récurrence sur n : $Q_n(t) = (e^{\lambda t} - 1)^{n-1}$, d'où $P_n(t) = e^{-\lambda t}(1 - e^{-\lambda t})^{n-1}$ pour $n \geq 1$. Comme $P_0(t) = 0$, la fonction génératrice de $N(t)$ est égale à $G_{1,t}(s) = \sum_{n\geq 1} e^{-\lambda t}(1 - e^{-\lambda t})^{n-1}s^n = se^{-\lambda t}(1 - s(1 - e^{-\lambda t}))^{-1}$. Si $n_0 \geq 2$, on peut supposer que le processus est régi par la somme de n_0 tels processus, ayant chacun un seul top initial en $t = 0$ et qui sont indépendants. La fonction génératrice est alors $(G_{1,t}(s))^{n_0}$.

4.7 **(a)** Comme la densité de U est $(1/t)I_{[0,t]}$, on a : $\mathrm{P}\{U + X \geq t\} = (1/t)\int_0^t \mathrm{P}\{U + X \geq t \,|\, U = s\}\,ds = (1/t)\int_0^t \mathrm{P}\{X \geq t - s \,|\, U = s\}\,ds = (1/t)\int_0^t \mathrm{P}\{X \geq t - s\}\,ds = (1/t)\int_0^t r(t-s)\,ds = (1/t)\int_0^t r(u)\,du = \overline{r}(t)$.

(b) Pour $k \geq 0$, posons $Y_k := I_{\{S_k + X_k \geq t\}}$. Comme $Q(t) = \sum_{k=0}^{N(t)} Y_k$, on a : $h(u) = \mathbb{E}[u^{Q(t)}] = \sum_{n\geq 0} \mathrm{P}\{N(t) = n\}\,\mathbb{E}[u^{Q(t)} \,|\, N(t) = n] = \sum_{n\geq 0} \mathrm{P}\{N(t) = n\} \times \mathbb{E}[u^{Y_0 + \cdots + Y_n} \,|\, N(t) = n] = \sum_{n\geq 0} \mathrm{P}\{N(t) = n\}\mathbb{E}[\prod_{k=0}^{n} u^{Y_k} \,|\, N(t) = n]$. Or, conditionnellement à $\{N(t) = n\}$ $(n \geq 1)$, le système $(S_1, \ldots, S_n)$ est distribué comme l'échantillon ordonné d'un système $(U_1, \ldots, U_n)$ de variables aléatoires indépendantes, uniformément réparties sur $[0, t]$. Pour $1 \leq k \leq n$, posons $Z_k := I_{\{U_k + X_k \geq t\}}$. On obtient : $h(u) = \sum_{n\geq 0} \mathrm{P}\{N(t) = n\}\,\mathbb{E}[\prod_{k=1}^{n} u^{Z_k}] = \sum_{n\geq 0} \mathrm{P}\{N(t) = n\}\,(\mathbb{E}[u^{Z_1}])^n$. Or Z_1 est une variable aléatoire de Bernoulli, de paramètre $\mathrm{P}\{U_1 + X_1 \geq t\} = \overline{r}(t)$. Donc $\mathbb{E}[u^{Z_1}] = 1 - \overline{r}(t) + u\,\overline{r}(t)$ et $h(u) = \sum_{n\geq 0} (e^{-\lambda t}(\lambda t)^n/n!)(1 - \overline{r}(t) + u\,\overline{r}(t))^n = e^{-\lambda t\,\overline{r}(t)(1-u)}$.

(c) On peut encore écrire $h(t) = e^{-m(t)(1-u)}$, ce qui montre que $Q(t)$ suit une loi de Poisson de paramètre $m(t)$.

Remarque : Lorsque t tend vers l'infini, on a $Q(t) \xrightarrow{\mathcal{L}} V$, où V est une variable aléatoire de Poisson, de paramètre $\lambda\,\mathbb{E}[X]$. En effet, lorsque t tend vers l'infini, on a : $m(t) = \lambda\int_0^t r(u)\,du \to \lambda\int_0^{+\infty} r(u)\,du = \lambda\,\mathbb{E}[X]$.

4.8 **(a)** D'abord, $\mathrm{P}\{N(t+h) - N(t) = 0 \,|\, N(t) = n\} = 1 - \mathrm{P}\{N(t+h) - N(t) \geq 1 \,|\, N(t) = n\} - \mathrm{P}\{N(t+h) - N(t) \leq -1 \,|\, N(t) = n\} = 1 - \lambda_n h + \mu_n h + o(h)$. Ensuite, pour $n \geq 1$, on trouve $\mathrm{P}\{N(t+h) = n\} = (1 - \lambda_n h - \mu_n h)P_n(t) + \lambda_{n-1}hP_{n-1}(t) + \mu_{n+1}hP_{n+1}(t) + o(h)$, puis $\mathrm{P}\{N(t+h) = 0\} = (1 - \lambda_0 h)P_0(t) + \mu_1 hP_1(t) + o(h)$.

(b) En faisant tendre h vers 0, on trouve bien le système différentiel annoncé.

(c) Avec $\lambda_n = n\lambda$ et $\mu_n = n\mu$, le système différentiel se spécialise en : $P'_n(t) = -n(\lambda+\mu)P_n(t) + (n-1)\lambda P_{n-1}(t) + (n+1)\mu P_{n+1}(t)$ $(n \geq 1)$; $P'_0(t) = \mu P_1(t)$; $P_n(0) = \delta_{n,n_0}$, ce qui conduit à l'équation aux dérivées partielles annoncée pour la fonction génératrice de $N(t)$. Sa résolution peut se faire à l'aide de la technique des intégrales premières.

CHAPITRE 5

5.1 **(a)** En dehors de la valeur propre 1, la matrice $\mathcal{P}$ a la valeur propre $\lambda \neq 1$ solution de $\det(\mathcal{P} - \lambda I) = 0$, soit $\lambda^2 - \lambda(2-\alpha-\beta) + 1 - \alpha - \beta = 0$, c'est-à-dire $\lambda = 1 - \alpha - \beta$, auquel on peut associer le vecteur propre $(\alpha, -\beta)$, d'où $S = \begin{pmatrix} 1 & \alpha \\ 1 & -\beta \end{pmatrix}$ et $S^{-1} = \dfrac{1}{\alpha+\beta}\begin{pmatrix} \beta & \alpha \\ 1 & -1 \end{pmatrix}$.

(b) $$\mathcal{P}^n = \begin{pmatrix} 1 & \alpha \\ 1 & -\beta \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & (1-\alpha-\beta)^n \end{pmatrix}\frac{1}{\alpha+\beta}\begin{pmatrix} \beta & \alpha \\ 1 & -1 \end{pmatrix} = \frac{1}{\alpha+\beta}\begin{pmatrix} \beta & \alpha \\ \beta & \alpha \end{pmatrix} + \frac{(1-\alpha-\beta)^n}{\alpha+\beta}\begin{pmatrix} \alpha & -\alpha \\ -\beta & \beta \end{pmatrix}$$ et $\lim_n \mathcal{P}^n = \dfrac{1}{\alpha+\beta}\begin{pmatrix} \beta & \alpha \\ \beta & \alpha \end{pmatrix}$.

5.2 **(a)** On a $\det(I - u\mathcal{P}) = (1-u)^2 + (\alpha+\beta)u(1-u)$ et
$$(I - u\,\mathcal{P})^{-1} = \frac{1}{(1-u)^2 + (\alpha+\beta)u(1-u)}\begin{pmatrix} 1-u+u\beta & u\beta \\ u\alpha & 1-u+u\alpha \end{pmatrix} := \begin{pmatrix} p_{1,1}(u) & p_{1,2}(u) \\ p_{2,1}(u) & p_{2,2}(u) \end{pmatrix}.$$

(b) Après quelques manipulations, on peut obtenir
$p_{1,1}(u) = \dfrac{\beta}{\alpha+\beta} + \dfrac{\alpha}{\alpha+\beta}(1-\alpha-\beta)^n$ et
$p_{2,1}(u) = \dfrac{\beta}{\alpha+\beta}\sum_{n\geq 0} u^n - \dfrac{\beta}{\alpha+\beta}(1-\alpha-\beta)^n$ et des expressions analogues pour $p_{2,2}(u)$ et $p_{1,2}(u)$ et retrouver les expressions de l'Exercice 5.1.

5.3 **(a)** D'abord, 1 est valeur propre et $\mathcal{P} - (\alpha-1)\,I = \begin{pmatrix} 1-\alpha & 1-\alpha & \alpha \\ 1-\alpha & 1-\alpha & \alpha \\ 1/3 & 1/3 & 4/3-\alpha \end{pmatrix}$ est singulière, ainsi que $\mathcal{P} - (\frac{1}{3} - \alpha)\,I = \begin{pmatrix} -1/3+\alpha & 1-\alpha & \alpha \\ 1-\alpha & -1/3+\alpha & \alpha \\ 1/3 & 1/3 & \alpha \end{pmatrix}$, puisque la somme des deux premières colonnes, multipliée par α, est égale à la troisième colonne multipliée par $\frac{2}{3}$.

(b) On vérifie chaque fois l'identité $\mathcal{P} \cdot v = \lambda\, v$, lorsque v est le vecteur proposé (écrit comme vecteur-colonne) pour la valeur propre λ.

(c) On a : $S\,S^{-1} = I$.

(d) Avec la matrice-diagonale $D := \mathrm{diag}(1, \alpha - 1, 1/3 - \alpha)$, on a : $\mathcal{P}^n = S\,D^n\,S^{-1}$. Posant $\beta := \alpha - 1$ et $\gamma = \frac{1}{3} - \alpha$, $\delta := 2 + 3\alpha$, on calcule

explicitement :

$$2\delta\mathcal{P}^n = \begin{pmatrix} 2+\delta\beta^n+3\alpha\gamma^n & 2-\delta\beta^n+3\alpha\gamma^n & 6\alpha-6\alpha\gamma^n \\ 2-\delta\beta^n+3\alpha\gamma^n & 2+\delta\beta^n+3\alpha\gamma^n & 6\alpha-6\alpha\gamma^n \\ 2-2\gamma^n & 2-2\gamma^n & 6\alpha+4\gamma^n \end{pmatrix}.$$

Comme $\lim_n \beta^n = \lim_n \gamma^n = 0$, on trouve $\lim_n \mathcal{P}^n = \dfrac{1}{\delta}\begin{pmatrix} 1 & 1 & 3\alpha \\ 1 & 1 & 3\alpha \\ 1 & 1 & 3\alpha \end{pmatrix}$.On a bien $u\mathcal{P} = u$, avec $u = (1/\delta, 1/\delta, 3\alpha/\delta)$.

5.4 Soit $E_0 := f^{n-1}(E)$; si $i \in E_0$, alors $f(i) = i$; réciproquement, cette relation entraîne $i = f(i) = f^2(i) = \cdots = f^{n-1}(i)$, d'où $i \in E_0$. Par conséquent, pour tout $j \in E$, il existe un point fixe i de f tel que $f^{n-1}(j) = i$. Toutes les matrices $\mathcal{P}_{f^m}$ sont identiques pour $m \geq n-1$. Soit $(a_{i,j})$ $(1 \leq i, j \leq n)$ la matrice $\mathcal{P}_{f^{n-1}}$. Sa colonne d'indice j est nulle si $j \notin E_0$; si $j \in E_0$, on a $a_{i,j} = 1$ si $f^{n-1}(i) = j$ et 0 autrement.

Le graphe associé à l'exemple et la matrice limite sont les suivants :

$$\lim_n \mathcal{P}_{f^n} = \begin{pmatrix} 0&0&0&0&0&0&1&0&0 \\ 0&0&0&0&0&0&1&0&0 \\ 0&0&0&0&0&0&1&0&0 \\ 0&0&0&0&0&1&0&0&0 \\ 0&0&0&0&0&0&1&0&0 \\ 0&0&0&0&0&1&0&0&0 \\ 0&0&0&0&0&0&1&0&0 \\ 0&0&0&0&0&0&1&0&0 \\ 0&0&0&0&0&1&0&0&0 \end{pmatrix}.$$

5.5 **(a)** En effet, on a : $\mu_i^{(n)} = \mathrm{P}\{X_n = i\} = \sum\limits_{k\in E} \mathrm{P}\{X_0 = k, X_n = i\} = \sum\limits_{k\in E} \mathrm{P}\{X_0 = k\}\,\mathrm{P}\{X_n = i \,|\, X_0 = k\} = \sum\limits_{k\in E} \mu_k^{(0)}\, p_{k,i}^{(n)}$.

(b) $\mathrm{P}\{X_m = i, X_n = j\} = \mathrm{P}\{X_m = i\}\,\mathrm{P}\{X_n = j \,|\, X_m = i\} = \mathrm{P}\{X_m = i\}\mathrm{P}\{X_{n-m} = j \,|\, X_0 = i\} = \mu_i^{(m)}\, p_{i,j}^{(n-m)}$.

5.6 **(a)** Tous les états communiquent. C'est clair !

(b) Un nombre *pair* de transitions est nécesaire pour aller de l'état 0 à l'état 0, donc $p_{0,0}^{(2n+1)} = 0$. Pour évaluer $p_{0,0}^{(2n)}$, on note que les chemins qui conviennent doivent avoir n transitions à gauche et n transitions à droite. Leur nombre est évidemment $\binom{2n}{n}$ et ils ont tous la probabilité $p^n(1-p)^n$, d'où la formule pour $p_{0,0}^{(2n)}$.

(c) Avec $n! \sim (n/e)^n\sqrt{2\,\pi\,n}$, on obtient :

$$p_{0,0}^{(2n)} \sim \frac{(2n/e)^{2n}\sqrt{2\,\pi\,2n}}{(n/e)^{2n}\,2\,\pi\,n} p^n(1-p)^n = \frac{(4p(1-p))^n}{\sqrt{\pi\,n}}.$$

Les séries de terme général $p_{0,0}^{(2n)}$ et $(4p(1-p))^n/\sqrt{\pi\,n}$ ont même comportement. Or $4p(1-p) = 1$ si $p = \frac{1}{2}$ et $4p(1-p) < 1$ si $p \neq \frac{1}{2}$. Par conséquent, la série

de terme général $p_{0,0}^{(2n)}$ diverge si $p = \frac{1}{2}$ et converge si $p \neq \frac{1}{2}$. On applique le critère de récurrence (Théorème 5.6.2).

(d) On a : $\binom{2n}{n} = 4^n\,(1/2)_n/n!$, où $(a)_n$ désigne la factorielle montante $a(a+1)\cdots(a+n-1)$ si $n \geq 1$ et 1 autrement. Il en résulte $P_{0,0}(s) = \sum_{n\geq 0}(4^n\,(1/2)_n/n!)p^nq^ns^{2n} = \sum_{n\geq 0}(1/2)_n\,(4pqs^2)^n = (1-4pqs^2)^{-1/2}$, par le théorème binomial.

(e) Par la formule (6.9), on a : $P_{0,0}(s) = (1-F_{0,0}(s))^{-1}$, d'où $F_{0,0}(s) = 1-(1-4pqs^2)^{1/2}$ et $F_{0,0}(1) = 1$ si $p = \frac{1}{2}$ et $F_{0,0}(1) < 1$ si $p \neq \frac{1}{2}$. L'état 0 (ainsi que tout autre état) est donc seulement récurrent lorsque $p = \frac{1}{2}$ et transient autrement.

5.7 Pour chacune des d composantes, on a $p_{0,0}^{(2n)} \sim 1/\sqrt{\pi\,n}$, d'après l'Exercice 5.6 ; d'où $p_{(0,\dots,0),(0,\dots,0)}^{(2n)} \sim (1/\sqrt{\pi\,n})^d$, qui est le terme général d'une série qui diverge si $d \leq 2$ (récurrence) et qui converge si $d > 2$ (transience).

5.8 Il suffit de montrer que $\mathrm{P}\{X_{n+1} = 1 \,|\, X_n = 0, X_{n-1} = 1\}$ est différent de $\mathrm{P}\{X_{n+1} = 1 \,|\, X_n = 0\}$. Or, l'évènement $\{X_{n+1} = 1, X_n = 0, X_{n-1} = 1\} = \{Y_{n+1} + Y_{n+2} = 1, Y_n$ et Y_{n+1} de signe contraire, $Y_{n-1} = Y_{n+1} = 1\}$ est vide, donc de probabilité nulle. Par conséquent, $\mathrm{P}\{X_{n+1} = 1 \,|\, X_n = 0, X_{n-1} = 1\} = 0$. Au contraire, $\mathrm{P}\{X_{n+1} = 1 \,|\, X_n = 0\} = \mathrm{P}\{X_{n+1} = 1, X_n = 0\}/\mathrm{P}\{X_n = 0\} = \mathrm{P}\{Y_{n+2} = Y_{n+1} = 1, Y_n = -1\}/\mathrm{P}\{Y_n$ et Y_{n+1} de signe contraire$\} = p^2q/(2pq) = p/2 > 0$.

5.9 **(a)** Évident.

(b) Le calcul de $\mathrm{P}\{X_{n+1} = j \,|\, X_n = i\}$ dépend du signe de $D_n - C_n = -X_n + 3 - Y_n$. Pour $D_n - C_n > 0$, soit $Y_n < 3 - X_n$, on obtient :

$i_1 = 1,\ j = 2,\quad \mathrm{P}\{Y_n < 2\} = 1,\quad p_{1,2} = 1$;
$i_1 = 2,\ j = 3,\quad \mathrm{P}\{Y_n < 1\} = 1/2,\quad p_{2,3} = 1/2$;
$i_1 = 3,\ j = 4,\quad \mathrm{P}\{Y_n < 0\} = 1/2,\quad p_{3,4} = 1/2$;
$i_1 = 4,\ j = 5,\quad \mathrm{P}\{Y_n < -1\} = 0,\quad p_{4,5} = 0$.

Pour $D_n - C_n < 0$, soit $Y_n > 3 - X_n$, on obtient, de façon analogue, $p_{1,0} = 0$, $p_{2,1} = 0$, $p_{3,2} = 1/2$, $p_{4,3} = 1/2$, $p_{5,4} = 1$. Enfin, pour $D_n - C_n = 0$, soit $Y_n = 3 - X_n$, on trouve $p_{1,1} = 0$, $p_{2,2} = 1/2$, $p_{3,3} = 0$, $p_{4,4} = 1/2$, $p_{5,5} = 0$. On en déduit la matrice de transition

$$\mathcal{P} = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array}\begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 1/2 & 1/2 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 1/2 & 1/2 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix} \end{array}.$$

Il y a trois classes indécomposables, les classes transientes $\{1\}$, $\{5\}$ et une classe récurrente $\{2,3,4\}$.

CHAPITRE 6

6.1 **(a)** Se reportant à (6.3.4) avec $a = 4$ et en notant φ_i la probabilité d'atteindre a_4 à partir de a_i, on trouve : $\varphi = (1-(q/p)^i)/(1-(q/p)^4$, lorsque $p \neq \frac{1}{2}$ et $i/4$, lorsque $p = \frac{1}{2}$ $(i = 1, 2, 3)$.

(b) Par un simple comptage, la probabilité que A gagne en arrivant au score $(50, 0)$ (resp. $(50, 15)$) est p^4 (resp. $4p^4q$).

(c) Pour entrer dans l'ensemble clos $\{a_0, a_1, a_2, a_3, a_4\}$ par $a_1 = (30, 40)$, il faut être passé en $(15, 40)$ dans l'étape précédente. D'où $p_1 = 4pq^3p$. Puis, $p_2 = \binom{4}{2}p^2q^2$ et $p_3 = 4qp^3q$.

(d) D'abord, l'évènement $A_1 := \{T_1 < \min\{T_2, T_3\}\}$ entraîne $\{X_5 = (30, 40)\}$ et $A_3 := \{T_3 < \min\{T_1, T_2\}\} \Rightarrow \{X_5 = (40, 30)\}$. De plus, ils sont antérieurs à 5. De même, $A_2 := \{T_2 < \min\{T_1, T_3\}\} \Rightarrow \{X_4 = (30, 30)\}$ et A_2 est antérieur à 4. Soit A_4 l'évènement « le score atteint $(50, 30)$ ». Notons P la probabilité de la chaîne à partir du score $(0, 0)$. On a : $\mathrm{P}(A_4) = \mathrm{P}(A_4 A_1) + \mathrm{P}(A_4 A_2) + \mathrm{P}(A_4 A_3) = \mathrm{P}(A_4 \mid A_1)\mathrm{P}(A_1) + \mathrm{P}(A_4 \mid A_2)\mathrm{P}(A_2) + \mathrm{P}(A_4 \mid A_3)\mathrm{P}(A_3)$. Or, il vient : $\mathrm{P}(A_4 \mid A_1) = \mathrm{P}\{A_4 \mid X_5 = (40, 30), A_1\} = \mathrm{P}\{A_4 \mid X_5 = (40, 30)\}$ et comme A_4 est postérieur à 5, la chaîne a atteint l'ensemble clos. On obtient : $\mathrm{P}(A_4 \mid A_1) = \varphi_1$. De même, $\mathrm{P}(A_4 \mid A_2) = \varphi_2$ et $\mathrm{P}(A_4 \mid A_3) = \varphi_3$. D'où $\mathrm{P}(A_4) = \varphi_1 p_1 + \varphi_2 p_2 + \varphi_3 p_3 = (10p^4 - 20p^5 + 10p^6)/(1 - 2p + 2p^2)$, une quantité qu'on ajoute à $p^4 + 4p^4q$ pour trouver la probabilité que A gagne le jeu, lorsque $p \neq \frac{1}{2}$. Notons que lorsque $p = \frac{1}{2}$, on trouve $\frac{1}{2}$ pour la probabilité que A (ou B) gagne le jeu. C'est aussi la valeur de $F(p) = \frac{1}{2}$, pour $p = \frac{1}{2}$. L'identité $F(p) + F(1-p) = 1$ signifie que le jeu se termine presque sûrement en un temps fini.

6.2 **(a)** $\mathcal{P} = \begin{pmatrix} 1/2 & 1/4 & 1/4 \\ 0 & 1/2 & 1/2 \\ 0 & 1/2 & 1/2 \end{pmatrix}$ $(r = 2)$ et $\mathcal{P} = \begin{pmatrix} 1/2 & 1/6 & 1/6 & 1/6 \\ 0 & 1/2 & 1/2 & 0 \\ 0 & 0 & 1/2 & 1/2 \\ 0 & 1/2 & 0 & 1/2 \end{pmatrix}$ $(r = 3)$. Il y a une classe transiente $\{0\}$ et une classe récurrente $\{1, 2, \ldots, r\}$. La chaîne est apériodique.

(b) Pour $r = 2$, le polynôme caractéristique de $\mathcal{P}$ a les racines 1, $\frac{1}{2}$ et 0, d'où $\mathcal{P} = \begin{pmatrix} 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \\ -1/2 & 0 & 1/2 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1/2 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 & -1 \\ 2 & -1 & -1 \\ 0 & 1 & 1 \end{pmatrix}$. On en tire : $\mathcal{P}^n = \begin{pmatrix} 1/2^n & 1/2 - 1/2^{n+1} & 1/2 - 1/2^{n+1} \\ 0 & 1/2 & 1/2 \\ 0 & 1/2 & 1/2 \end{pmatrix}$ et $\begin{pmatrix} 0 & 1/2 & 1/2 \\ 0 & 1/2 & 1/2 \\ 0 & 1/2 & 1/2 \end{pmatrix}$ pour la valeur limite $\lim_n \mathcal{P}^n$.

(c) Pour déterminer $f_{1,1}^{(n)}$, il suffit de compter les chemins allant de 1 à 1 en n transitions, sans passer par 1. Or, pour un tel chemin, la première transition est $1 \rightsquigarrow 2$ et la dernière $r \rightsquigarrow 1$. Un tel chemin est caractérisé par la suite $(i_2, i_3, \ldots, i_{r-1}, i_r)$ des instants où le processus se trouve pour la dernière fois en $2, 3, \ldots, (r-1), r$. Comme $1 \leq i_2 < i_3 < \cdots < i_{r-1} < i_r = (n-1)$, il y a

$\binom{n-2}{r-2}$ tels chemins de longueur n et $f_{1,1}^{(n)} = \binom{n-2}{r-2}(1/2)^n$ $(n \geq r)$ et $f_{1,1}^{(1)} = 1/2$. La fonction génératrice de T_1 vaut

$$G(s) = \frac{s}{2} + \sum_{n \geq r} \binom{n-2}{r-2} \left(\frac{s}{2}\right)^n = \frac{s}{2} + \left(\frac{s}{2}\right)^r \left(1 - \frac{s}{2}\right)^{-r+1},$$

d'où l'on tire $\mathbb{E}[T_1] = G'(1) = r$, $\mathbb{E}[T_1(T_1 - 1)] = G''(1) = 2r(r-1)$ et $\operatorname{Var} T_1 = r(r-1)$.

(d) Comme 0 est transient, pour tout état i, on a : $\lim_n p_{i,0}^{(n)} = 0$. D'autre part, la chaîne n'a qu'une seule classe récurrente. D'après le Théorème 4.4, il existe une loi stationnaire $(\pi_0, \pi_1, \ldots, \pi_r)$ unique telle que $\pi_j = \sum_i \pi_i p_{i,j}$ $(0 \leq j \leq r)$, un système qui se récrit : $\pi_0 = 0$, $\pi_1 = \frac{1}{2}(\pi_1 + \pi_r)$, $\pi_j = \frac{1}{2}(\pi_{j-1} + \pi_j)$ $(2 \leq j \leq r)$. D'où $\pi_0 = 0$ et $\pi_j = 1/r$ $(1 \leq j \leq r)$. Maintenant, l'unique classe récurrente est apériodique. On peut donc appliquer le Théorème 6.1 et conclure que, pour $1 \leq j \leq r$, on a $p_{i,j}^{(n)} \to \pi_j = 1/r$ lorsque n tend vers l'infini.

6.3 Tous les états communiquent. Le processus est donc *irréductible.* Les états sont ou apériodiques, ou périodiques de période 2, selon que r est impair ou pair. Le processus est donc ergodique si et seulement si r est impair. La matrice $\mathcal{P}$ est *bistochastique.* D'après le Théorème 6.7.1, le processus admet une et une seule loi de probabilité stationnaire, la loi uniformément répartie $(1/r, \ldots, 1/r)$. La fréquence asymptotique moyenne de séjour dans chaque état est $1/r$ et le temps moyen de retour dans chaque état est r. Il est à noter que ces résultats sont *indépendants* de p $(0 < p < 1)$.

6.4 **(a)** La matrice de transition est donnée par la matrice $\mathcal{P}$ ci-après.

$$\mathcal{P} = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \end{array} \begin{array}{c} \begin{array}{ccccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \end{array} \\ \begin{pmatrix} 0 & 1/2 & 0 & 0 & 0 & 1/2 & 0 & 0 & 0 \\ 1/3 & 0 & 1/3 & 0 & 1/3 & 0 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/3 & 0 & 1/3 & 0 & 0 & 0 & 1/3 \\ 0 & 1/4 & 0 & 1/4 & 0 & 1/4 & 0 & 1/4 & 0 \\ 1/3 & 0 & 0 & 0 & 1/3 & 0 & 1/3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 0 & 1/3 & 0 & 1/3 & 0 & 1/3 \\ 0 & 0 & 0 & 1/2 & 0 & 0 & 0 & 1/2 & 0 \end{pmatrix} \end{array}.$$

(b) Tous les états communiquent et sont de période 2. Le processus est donc *irréductible*, mais il n'est *pas ergodique.*

(c) Il y a une et une seule loi de probabilité stationnaire. La résolution du système $u\mathcal{P} = u$ semble inextricable. Or, on sait que u_j est la fréquence asymptotique moyenne de séjour dans le compartiment j. On peut imaginer que ce temps de séjour moyen dans un compartiment est proportionnel au nombre de portes de ce compartiment, autrement dit, que les u_j sont

proportionnels aux composantes du vecteur $(2,3,2,3,4,3,2,3,2)$. En effet, on vérifie facilement que

$$u = \left(\frac{2}{24}, \frac{3}{24}, \frac{2}{24}, \frac{3}{24}, \frac{4}{24}, \frac{3}{24}, \frac{2}{24}, \frac{3}{24}, \frac{2}{24}\right)$$

est solution de $u\mathcal{P} = u$. C'est donc la seule loi de probabilité stationnaire de la chaîne. Notons que les temps de séjour moyen dans les compartiments sont donnés par les inverses des composantes du vecteur u précédent.

6.5 Pour tout $n \geq 0$, désignons par X_n la distance qui sépare les deux coccinelles à la date n, ou encore le nombre minimum d'arêtes de l'hexagone entre les deux sommets où elles se trouvent. La suite (X_n) $(n \geq 0)$ est une chaîne de Markov homogène, dont l'espace des états est $E = \{0,1,2,3\}$ et la matrice de transition $\mathcal{P}$, ainsi que le graphe associé sont donnés ci-après :

$$\mathcal{P} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} \\ \begin{pmatrix} 1/2 & 0 & 1/2 & 0 \\ 0 & 3/4 & 0 & 1/4 \\ 1/4 & 0 & 3/4 & 0 \\ 0 & 1/2 & 0 & 1/2 \end{pmatrix} \end{array}$$

0 —1/2→ 2 ; 2 —1/4→ 0 ; boucles : 0 : 1/2, 2 : 3/4
1 —1/4→ 3 ; 3 —1/2→ 1 ; boucles : 1 : 3/4, 3 : 1/2

Il y a deux classes indécomposables $\{0,2\}$, $\{1,3\}$, toutes deux récurrentes. Désignons par d la distance initiale qui sépare les deux coccinelles. Il s'agit de calculer le temps moyen qu'elles mettront pour se rencontrer *pour la première fois*, soit $M_{d,0}$. Les relations $M_{1,0} = M_{3,0} = +\infty$ sont immédiates. Pour déterminer $M_{0,0}$ et $M_{2,0}$, il suffit de considérer la restriction du processus à l'ensemble $\{0,2\}$, dont la matrice de transition est $\begin{pmatrix} 1/2 & 1/2 \\ 1/4 & 3/4 \end{pmatrix}$. Ce processus restreint, qui a un nombre fini d'états, est irréductible; donc, pour tout i,j, on a : $M_{i,j} < +\infty$. Or, les $M_{i,j}$ vérifie le système : $M_{i,j} = 1 + \sum_{k \neq j} p_{i,k} M_{k,j}$ $(i,j \in \{0,2\})$. En particulier, $M_{0,0} = 1 + \frac{1}{2}M_{2,0}$ et $M_{2,0} = 1 + \frac{3}{4}M_{2,0}$, d'où $M_{0,0} = 3$, $M_{2,0} = 4$.

6.6 **(a)** Le processus (Y_n) $(n \geq 1)$ est de Markov. En effet, supposons que l'évènement $\{Y_{n-1} = (i_1,i_2)\}$ soit réalisé. Alors, comme le processus (X_n) $(n \geq 0)$ est de Markov, la loi de probabilité de Y_n, conditionnellement à Y_{n-1} est déterminée. On a :

$$\mathrm{P}\{Y_n = (j_1,j_2) \,|\, Y_{n-1} = (i_1,i_2)\} = \begin{cases} 0, & \text{si } i_2 \neq j_1 ; \\ p_{j_1,j_2}, & \text{si } i_2 = j_1. \end{cases}$$

(b) On a : $\mathrm{P}\{Y_n = (i,j)\} = \mathrm{P}\{X_{n-1} = i\}\,\mathrm{P}\{X_n = j \,|\, X_{n-1} = i\} = \mathrm{P}\{X_{n-1} = i\}\, p_{i,j}$. Comme le processus (X_n) $(n \geq 0)$ est ergodique, il existe une et une seule loi de probabilité stationnaire $u = (u_k)$, on en déduit : $\lim_n \mathrm{P}\{Y_n = (i,j)\} = u_j\, p_{i,j}$.

6.7 Comme on est dans le cas fini, il existe *au moins* une loi de probabilité stationnaire u. Elle vérifie $u = u\mathcal{P}$, ou encore $u_k = \sum_{i=0}^{a} u_i\, p_{i,k}$

$(i = 0, 1, \ldots, a)$, ce qui donne le système : $u_0 = u_0 + qu_1$, $u_1 = qu_2$, $\ldots$, $u_k = pu_{k-1} + qu_{k+1}$ $(2 \le k \le a-2)$, $\ldots$, $u_{a-1} = pu_{a-2}$, $u_a = pu_{a-1} + u_a$. La première équation donne $u_1 = 0$, d'où $u_1 = \cdots = u_{a-1} = 0$. La première et la dernière équations donnent $u_0 = u_0$ et $u_a = u_a$. Les quantités u_0, u_a sont donc indéterminées. Les quantités u_0 et u_a sont indéterminées. Pour tout α tel que $0 < \alpha < 1$, le vecteur $u := (\alpha, 0, \ldots, 0, 1-\alpha)$ est une loi de probabilité stationnaire. Il n'y a pas unicité (il existe, en effet, deux classes récurrentes). On constate également que toutes les lois stationnaires sont portées par les classes récurrentes $\{0\}$, $\{a\}$.

6.8 On considère un ensemble de $(m+1)$ états $0, 1, \ldots, m$, symbolisant les nombres de vignettes collées dans l'album. Après le premier achat, on passe de l'état 0 à l'état 1, avec une probabilité égale à 1. Si l'on est dans l'état 1, on peut y rester, avec une probabilité àgale à $1/m$ (on rachète la vignette que l'on a déjà), ou passer dans l'état 2, avec une probabilité égale à $(m-1)/m$, et ainsi de suite. Il s'agit donc de considérer une chaîne de Markov homogène dont la matrise de transition est donnée par :

$$\begin{array}{c} \\ 0 \\ 1 \\ 2 \\ \vdots \\ (m-2) \\ (m-1) \\ m \end{array}
\begin{array}{c} \begin{array}{ccccccc} 0 & 1 & 2 & \cdots & (m-2) & (m-1) & m \end{array} \\
\begin{pmatrix}
0 & 1 & 0 & \cdots & 0 & 0 & 0 \\
0 & 1/m & (m-1)/m & \cdots & 0 & 0 & 0 \\
0 & 0 & 2/m & \cdots & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
0 & 1 & 0 & \cdots & (m-2)/m & 2/m & 0 \\
0 & 1 & 0 & \cdots & 0 & (m-1)/m & 1/m \\
0 & 1 & 0 & \cdots & 0 & 0 & 1
\end{pmatrix} \end{array}$$

Dans les notations du Théorème 6.2.3, le temps moyen $\mathbb{E}[T]$ est le nombre $M_{0,m}$. Il s'agit de chercher à évaluer ce nombre à l'aide du système : $M_{i,m} = 1 + \sum_{k \ne m} p_{i,k} M_{k,m}$ $(i = 0, 1, \ldots, m-1)$, qui se récrit : $M_{0,m} = 1 + M_{1,m}$, $M_{1,m} = (m/(m-1)) + M_{2,m}$, $M_{2,m} = (m/(m-2)) + M_{3,m}$, $\ldots$, $M_{m-2,m} = (m/2) + M_{m-1,m}$, $M_{m-1,m} = m$. On en tire : $M_{0,m} = 1 + (m/(m-1)) + (m/(m-2)) + \cdots + m/2 + m = m\,H_m$.

6.9 **(a)** D'après la Règle 5, $\mathbb{E}^i[T] = \mathbb{E}^i[\mathbb{E}^i[T \mid X_1]] = \sum_{j \in \text{Rec}} p_{i,j}\, \mathbb{E}^i[T \mid X_1 = j] + \sum_{j \in \text{Tra}} p_{i,j}\, \mathbb{E}^i[T \mid X_1 = j] = \sum_{j \in \text{Rec}} p_{i,j} + \sum_{j \in \text{Tra}} p_{i,j}(1 + \mathbb{E}^j[T]) = 1 + \sum_{j \in \text{Tra}} p_{i,j} \mathbb{E}^j[T]$.

(b) D'après le Théorème 6.2.3, on obtient le système : $\mathbb{E}^1[T] = 1 + p\,\mathbb{E}^2[T]$; $\mathbb{E}^2[T] = 1 + p\mathbb{E}^3[T] + q\,\mathbb{E}^1[T]$; $\mathbb{E}^3[T] = 1 + q\,\mathbb{E}^2[T]$, qui a la solution unique $\mathbb{E}^1[T] = (1+2p^2)/(1-2pq)$; $\mathbb{E}^2[T] = 2/(1-2pq)$; $\mathbb{E}^3[T] = (1+2q^2)/(1-2pq)$.

(c) Soit T le temps d'atteinte dans l'un des états de fin de jeu. Alors $\mathbb{E}^{(40,15)}[T] = 1 + q\,\mathbb{E}^{(40,30)}[T]$. Or, $\mathbb{E}^{(40,30)}[T]$ est précisément égal à $\mathbb{E}^3[T]$ de (b). Par suite, $\mathbb{E}^{(40,15)}[T] = 1 + q(1+2q^2)/(1-2pq) = (4 - 9p + 8p^2 - 2p^3)/(1 - 2p + 2p^2)$.

(d) Directement sur le graphe de la chaîne, on calcule les probabilités $\mathrm{P}^{(40,15)}\{T=n\}$ et on trouve pour fonction génératrice : $G_T(s) = ps + pqs^2 + q^2(p^2+q^2)s^4/(1-2pqs^2)$, d'où l'on tire encore la valeur de $\mathbb{E}^{(40,15)}[T] = G'_T(1)$.

6.10 Considérons un vecteur propre $(x_1,\dots,x_r)$ associé à la valeur propre λ, de sorte que $(\lambda - p_{i,i})x_i = \sum_{j\neq i} p_{i,j}x_j \quad (i=1,\dots,r)$; en posant $K := \sup_{i=i}^{r} |x_i|$, on obtient $|\lambda - p_{i,i}|\,|x_i| \le K \sum_{j\neq i} p_{i,j} = K(1-p_{i,i})$ $(i=1,\dots,r)$. Soit h un indice satisfaisant $1 \le h \le r$ et $|x_h| = K > 0$. En écrivant la précédente inégalité pour $i=h$, on obtient : $|\lambda - p_{h,h}| \le 1 - p_{h,h}$. On en déduit : $|\lambda - \omega| \le |\lambda - p_{h,h}| + |p_{h,h} - \omega| \le (1-p_{h,h}) + (p_{h,h} - \omega) = 1 - \omega$.

CHAPITRE 7

7.1 On vérifie que les conditions (a) et (b) de la Définition 7.2.1 sont bien satisfaites. D'abord, $\mathbb{E}[|X_n|] \le 2n\sigma^2 < +\infty$; ensuite, on a $\mathbb{E}[X_{n+1} \mid Y_0,\dots,Y_n] = \mathbb{E}[(Y_{n+1} + \sum_{k=1}^{n} Y_k)^2 - (n+1)\sigma^2 \mid Y_0,\dots,Y_n] =$ $\mathbb{E}[Y_{n+1}^2 + 2Y_{n+1}\sum_{k=1}^{n} Y_k + (\sum_{k=1}^{n} Y_k)^2 - (n+1)\sigma^2 \mid Y_0,\dots,Y_n] =$ $\mathbb{E}[Y_{n+1}^2 + 2Y_{n+1}\sum_{k=1}^{n} Y_k + X_n - \sigma^2 \mid Y_0,\dots,Y_n]$. Or, $\mathbb{E}[Y_{n+1}^2 \mid Y_0,\dots,Y_n] = \mathbb{E}[Y_{n+1}^2] = \sigma^2$, car Y_{n+1}^2 est indépendant de $(Y_0,\dots,Y_n)$. On obtient ensuite : $\mathbb{E}[2Y_{n+1}\sum_{k=1}^{n} Y_k \mid Y_0,\dots,Y_n] = 2(\sum_{k=1}^{n} Y_k)\ \mathbb{E}[Y_{n+1} \mid Y_0,\dots,Y_n] = 2(\sum_{k=1}^{n} Y_k)$ $\times\ \mathbb{E}[Y_{n+1}] = 0$, en utilisant successivement la Règle 4, le fait que les Y_i sont indépendants et la Règle 1. Enfin, $\mathbb{E}[X_n - \sigma^2 \mid Y_0,\dots,Y_n] = X_n - \sigma^2$, d'après la Règle 2 et la Propriété 7.2.1. Ainsi, $\mathbb{E}[X_{n+1} \mid Y_0,\dots,Y_n] = X_n$.

7.2 D'abord $\mathbb{E}[|X_n|] < +\infty$ (il suffit, en fait, de supposer que f est bornée). Puis, $\mathbb{E}[X_{n+1} \mid Y_0,\dots,Y_n] = \mathbb{E}[f\circ Y_{n+1} \mid Y_0,\dots,Y_n] = \mathbb{E}[f\circ Y_{n+1} \mid Y_n]$, en vertu de la propriété de Markov. Ensuite, $\mathbb{E}[f\circ Y_{n+1} \mid Y_n = i] = \sum_j p_{ij} f(j)$. D'où $\mathbb{E}[f\circ Y_{n+1} \mid Y_n] = \sum_j p_{Y_n,j} f(j) = f\circ Y_n = X_n$, puisque $\mathcal{P}f = f$.

7.3 Voir solution du précédent exercice.

7.4 **(a)** On a $S_{n+1} = S_n + Y_{n+1}$ et S_n et Y_{n+1} sont indépendants. D'où $\mathrm{P}\{S_{n+1} = k_{n+1} \mid S_0 = k_0,\dots,S_n = k_n\} = \mathrm{P}\{S_{n+1} = k_{n+1} \mid S_n = k_n\}$. Puis, $p_{ij} = \mathrm{P}\{S_{n+1} = j \mid S_n = i\} = \mathrm{P}\{S_n + Y_{n+1} = j \mid S_n = i\} = \mathrm{P}\{Y_{n+1} = j - i \mid S_n = i\} = \mathrm{P}\{Y_{n+1} = j-i\} = q_{j-i}$, puisque Y_{n+1} et S_n sont indépendants.

(b) En effet,
$\sum_j p_{ij}e^{uj} = \sum_j q_{j-i}e^{uj} = \sum_k q_k e^{u(i+k)} = e^{ui}\sum_k q_k e^{uk} = e^{ui}g(u) = g(u)e^{ui}$.

(c) Il suffit d'appliquer le résultat de l'Exercice 7.3.

7.5 D'abord, $\mathbb{E}[X_n] = (\mathbb{E}[f_1 \circ Y_0 / f_0 \circ Y_0])^n = (\int (f_1(y)/f_0(y))\, f_0(y)\, dy)^n = (\int f_1(y)\, dy)^n = 1$. Ensuite,

$$\begin{aligned}\mathbb{E}[X_{n+1} \mid Y_0, \dots, Y_n] &= \mathbb{E}[X_n\, (f_1 \circ Y_{n+1} / f_0 \circ Y_{n+1}) \mid Y_0, \dots, Y_n] \\ &= X_n\, \mathbb{E}[(f_1 \circ Y_{n+1} / f_0 \circ Y_{n+1}) \mid Y_0, \dots, Y_n] \\ &= X_n\, \mathbb{E}[(f_1 \circ Y_{n+1} / f_0 \circ Y_{n+1})] \\ &\qquad\qquad [\text{puisque les } Y_i \text{ sont indépendants.}] \\ &= X_n \int (f_1(y)/f_0(y))\, f_0(y)\, dy = X_n \int f_1(y)\, dy \\ &= X_n.\end{aligned}$$

7.6 D'abord, $\mathbb{E}[|X_n|] = |c|\, \mathbb{E}[|Y_1| \dots |Y_n|] = |c|\, \mathbb{E}[|Y_1|] \cdots \mathbb{E}[|Y_n|] < +\infty$; puis, comme la variable X_n est une fonction des seules $Y_0, \dots, Y_n$, on a, d'après la Règle 4 et l'indépendance des Y_i : $\mathbb{E}[X_{n+1} \mid Y_0, \dots, Y_n] = \mathbb{E}[X_n\, Y_{n+1} \mid Y_0, \dots, Y_n] = X_n\, \mathbb{E}[Y_{n+1} \mid Y_0, \dots, Y_n] = X_n\, \mathbb{E}[Y_{n+1}] = X_n$.

7.7 Les variables aléatoires $e^{uY_n}/g(u)$ $(n \ge 1)$ sont indépendantes et ont même distribution et $X_n = (e^{uY_1}/g(u)) \cdots (e^{uY_n}/g(u))$.

7.8 On prend $\xi = \mathbb{E}[X]$ dans l'inégalité : $\varphi(x) \ge \lambda(x - \xi) + \varphi(\xi)$. Alors $\mathbb{E}[\varphi \circ X] = \int \varphi(x)\, d\mathrm{P}_X(x) \ge \lambda \int (x - \mathbb{E}[X])\, d\mathrm{P}_X(x) + \int \varphi(\mathbb{E}[X])\, d\mathrm{P}_X(x) = \varphi(\mathbb{E}[X])$. On prend ensuite $\xi = \mathbb{E}[X \mid Y = y]$, on multiplie l'inégalité initiale par la densité $f_{X \mid Y}(x \mid y)$, puis on intègre par rapport à la mesure dy.

7.9 Pour tout ≥ 1, on a : $\mathbb{E}[|X_n|] \le n\, \mathbb{E}[|Y_1|] < +\infty$. Puis, $\mathbb{E}[X_{n+1} \mid Y_1, \dots, Y_n] = \mathbb{E}[X_n + Y_{n+1} \mid Y_1, \dots, Y_n] = X_n + \mathbb{E}[Y_{n+1} \mid Y_1, \dots, Y_n]$ par linéarité et par la Règle 2, soit encore $X_n + \mathbb{E}[Y_{n+1}] = X_n + \mu$, puisque les Y_k sont indépendantes et de même loi, une quantité qui est $\ge X_n$, $\le X_n$, $= X_n$, suivant que $\mu \ge 0$, $\mu \le 0$, $\mu = 0$.

7.10 D'abord, $\mathbb{E}[|X_n|] = \sum_{1 \le k \le n} \mathbb{E}[Y_k^2] < +\infty$. Puis, $\mathbb{E}[X_{n+1} \mid Y_1, \dots, Y_n] = \mathbb{E}[(S_n + Y_{n+1})^2 \mid Y_1, \dots, Y_n] = \mathbb{E}[S_n^2 + 2S_n Y_{n+1} + Y_{n+1}^2 \mid Y_1, \dots, Y_n] = S_n^2 + 2S_n\, \mathbb{E}[Y_{n+1} \mid Y_1, \dots, Y_n] + \mathbb{E}[Y_{n+1}^2 \mid Y_1, \dots, Y_n]$ par les Règles 3 et 4. Ensuite, en vertu de l'indépendance des Y_k et du fait que les Y_k sont centrées, on obtient : $\mathbb{E}[X_{n+1} \mid Y_1, \dots, Y_n] = S_n^2 + 2S_n\, \mathbb{E}[Y_{n+1} \mid Y_1, \dots, Y_n] + \mathbb{E}[Y_{n+1}^2] = S_n^2 + \mathbb{E}[Y_{n+1}^2 \mid Y_1, \dots, Y_n] \ge S_n^2 = X_n$.
Autre démonstration : d'après la Proposition 7.2.5, la suite (S_n) $(n \ge 1)$ est une martingale par rapport à (Y_n) $(n \ge 1)$. Or $\varphi : x \mapsto x^2$ est convexe et $X_n = (S_n)^2 = \varphi \circ S_n$. Par ailleurs, $\mathbb{E}[(\varphi \circ S_n)^+] = \mathbb{E}[(S_n^2)^+] = \mathbb{E}[S_n^2] = \mathbb{E}[\operatorname{Var} S_n] = \sum_{k=1}^{n} \operatorname{Var} Y_k = \sum_{k=1}^{n} \mathbb{E}[Y_n^2] < +\infty$. D'après le Lemme 7.4.1, la suite (X_n) $(n \ge 1)$ est donc une sous-martingale par rapport à (Y_n) $(n \ge 1)$.

7.11 D'abord, $\mathbb{E}[|Y_n|] \le \dfrac{1}{n} \sum_{k=1}^{n} \mathbb{E}[|X_1|] < +\infty$. Ensuite, comme $Y_{n-1} = \dfrac{n}{n-1} Y_n - \dfrac{X_n}{n-1}$, on a $\mathbb{E}[Y_{n-1} \mid Y_n, Y_{n+1}, \dots] = \dfrac{n}{n-1} \mathbb{E}[Y_n \mid Y_n, Y_{n+1}, \dots] -$

$\dfrac{1}{n-1}\mathbb{E}[X_n \mid Y_n, Y_{n+1}, \dots] = \dfrac{n}{n-1}Y_n - \dfrac{1}{n-1}\mathbb{E}[X_n \mid Y_n, Y_{n+1}, \dots]$, d'après la Règle 4. Or, puisque les variables aléatoires $X_1, \dots, X_n$ sont indépendantes et identiquement distribuées, $\mathbb{E}[X_1 \mid Y_n, Y_{n+1}, \dots] = \dots = \mathbb{E}[X_n \mid Y_n, Y_{n+1}, \dots]$, d'où $Y_n = \mathbb{E}[Y_n \mid Y_n, Y_{n+1}, \dots] = \dfrac{1}{n}\sum_{k=1}^{n} \mathbb{E}[X_k \mid Y_n, Y_{n+1}, \dots] = \mathbb{E}[X_n \mid Y_n, Y_{n+1}, \dots]$, d'où, en définitive, $\mathbb{E}[Y_{n-1} \mid Y_n, Y_{n+1}, \dots] = \dfrac{n}{n-1}Y_n - \dfrac{1}{n-1}Y_n = Y_n$.

CHAPITRE 8

8.1 La comparaison entre $\mathbb{E}[X_0]$ et $\mathbb{E}[X_T]$ a été démontrée dans le Théorème 7.2.1. Il suffit de comparer seulement $\mathbb{E}[X_T]$ et $\mathbb{E}[X_M]$. On a $X_T = (\sum_{0\le k\le M} I_{\{T=k\}})X_T = \sum_{0\le k\le M} I_{\{T=k\}}X_T = \sum_{0\le k\le M} I_{\{T=k\}}X_k$, d'où $\mathbb{E}[X_T] = \sum_{0\le k\le M} \mathbb{E}[I_{\{T=k\}}X_k]$. De même, $\mathbb{E}[X_M] = \sum_{0\le k\le M} \mathbb{E}[I_{\{T=k\}}X_M] =$ $\sum_{0\le k\le M} \mathbb{E}[\mathbb{E}[I_{\{T=k\}}X_M \mid Y_0, \dots, Y_k]] = \sum_{0\le k\le M} \mathbb{E}[I_{\{T=k\}}\mathbb{E}[X_M \mid Y_0, \dots, Y_k]]$, car chaque indicatrice $I_{\{T=k\}}$ est une fonction de $(Y_0, \dots, Y_k)$. D'après la Propriété 7.4.4, on en tire $\mathbb{E}[X_M] = \sum_{0\le k\le M} \mathbb{E}[I_{\{T=k\}}X_k]$ (resp. $\ge \sum_{0\le k\le M} \mathbb{E}[I_{\{T=k\}}X_k]$, resp. $\le \sum_{0\le k\le M} \mathbb{E}[I_{\{T=k\}}X_k]$), si X est une martingale (resp. une sous-martingale, resp. une surmartingale).

La variable aléatoire $T\wedge n$ est un temps d'arrêt *borné* ($M = n$). Pour obtenir les relations pour le processus arrêté, il suffit donc, dans les relations précédentes, de remplacer T par $T \wedge n$ et M par n.

8.2 La variable aléatoire T est un temps d'arrêt du processus (Y_n) et suit évidemment la loi géométrique, de sorte que $\mathbb{E}[T] < +\infty$. On est dans les conditions d'application du Corollaire 8.3.4. Comme évidemment $X_T = 1$, on a : $1 = \mathbb{E}[X_T] = \mathbb{E}[T]\,\mathbb{E}[Y_1] = \mathbb{E}[T]\,p$, d'où $\mathbb{E}[T] = 1/p$. De même, $X_T = r$ et $r = \mathbb{E}[T]\,\mathbb{E}[Y_1] = \mathbb{E}[T]\,p$, d'où $\mathbb{E}[T] = r/p$.

8.3 **(A)** (a) On a toujours $X_T = 1$. Comme $p = q = \frac{1}{2}$, on a $\mathbb{E}[Y_n] = 0$ et aussi $\mathbb{E}[X_n] = 0$ pour $n \ge 1$. Si l'on avait $\mathbb{E}[T] < +\infty$, le Corollaire 7.3.4 permettrait d'écrire : $1 = \mathbb{E}[X_T] = \mathbb{E}[T]\,\mathbb{E}[Y_1] = 0$, d'où une contradiction. Ainsi $\mathbb{E}[T] = +\infty$.

(b) En prenant $U = X^{|T}$ et $V = 1$, on peut conclure que $\mathrm{P}\{T < +\infty\} = 1$. D'autre part, la martingale $X = (X_n)$ ne converge pas, puisque le terme général Y_n ne tend pas vers 0. Or, sur $\{T = +\infty\}$, les deux martingales X et $X^{|T}$ coïncident. L'ensemble $\{T = +\infty\}$ est donc négligeable.

(c) On a : $(1/(2n+1))\binom{2n+1}{n+1}(1/2^{2n+1}) = -(-1/2)_{n+1}/(n+1)!$ D'où $-\sum_{n\ge 0}(-1/2)_{n+1}/(n+1)! = (1-(1-u)^{1/2})\,|_{u\,=\,1} = 1$.

(B) On a : $1 = \mathbb{E}[Z_0] = \mathbb{E}[Z_{T\wedge n}] \geq \mathbb{E}[Z_{T\wedge n} I_{\{T\leq n\}}] = \mathbb{E}[Z_T I_{\{T\leq n\}}] = \mathbb{E}[(q/p)^{X_T} I_{\{T\leq n\}}]$. Puisque $X_T = 1$, on en tire : $1 \geq (q/p)\mathbb{E}[I_{\{T\leq n\}}] = (q/p)\mathrm{P}\{T \leq n\}$, d'où $\mathrm{P}\{T \leq n\} \leq p/q < 1$. En faisant tendre n vers l'infini, il vient $\mathrm{P}\{T < +\infty\} \leq p/q < 1$. Il en résulte que $\mathbb{E}[T] = +\infty$, un résultat que l'on peut démontrer directement : si l'on avait $\mathbb{E}[T] < +\infty$, d'après le Corollaire 8.3.4 ("les identités de Wald"), on en déduirait $1 = \mathbb{E}[X_T] = \mathbb{E}[T]\,\mathbb{E}[Y_1] = \mathbb{E}[T]\,(p-q)$, soit $\mathbb{E}[T] < 0$, puisque $p - q < 0$, d'où une contradiction.

(C) On a : $X_n/n = (Y_1 + \cdots + Y_n)/n \xrightarrow{p.s.} \mathbb{E}[Y_1] = p - q > 0$. D'où $X_n \xrightarrow{p.s.} +\infty$. Comme $T = \inf\{n \geq 1 : X_n = 1\}$, on a bien T presque sûrement fini. Par définition, $1 \geq X_{T\wedge n} = Y_1 + \cdots + Y_{T\wedge n}$. Comme $\mathbb{E}[T \wedge n] \leq n < +\infty$, on peut appliquer le Corollaire 3.4 : pour tout $n \geq 1$, on a : $1 \geq \mathbb{E}[X_{T\wedge n}] = \mathbb{E}[T \wedge n]\,\mathbb{E}[Y_1] = \mathbb{E}[T \wedge n](p - q)$ et donc $\mathbb{E}[T \wedge n] \leq 1/(p - q)$. Comme T est presque sûrement fini, on en déduit aussi que $T \wedge n \xrightarrow{p.s.} T$. D'après le lemme de Fatou, on peut aussi en conclure que $\mathbb{E}[T] \leq 1/(p - q)$. On applique de nouveau le Corollaire 8.3.4 à X_T : comme $\mathbb{E}[T] < +\infty$, on obtient : $1 = \mathbb{E}[X_T] = \mathbb{E}[T]\,\mathbb{E}[Y_1] = \mathbb{E}[T](p-q)$. D'où $\mathbb{E}[T] = 1/(p-q)$.

CHAPITRE 9

9.1 On a : $\mathbb{E}[\,|Z_n|\,] = \mathbb{E}[Z_n] = (\mathbb{E}[(q/p)^{Y_1}])^n = (p(q/p) + q(p/q))^n = 1$. Puis, $\mathbb{E}[Z_{n+1} \,|\, Y_0, \ldots, Y_n] = \mathbb{E}[Z_n(q/p)^{Y_{n+1}} \,|\, Y_0, \ldots, Y_n] = Z_n\mathbb{E}[(q/p)^{Y_{n+1}} \,|\, Y_0, \ldots, Y_n] = Z_n\mathbb{E}[(q/p)^{Y_{n+1}}] = Z_n(p(q/p) + q(p/q)) = Z_n$.

9.2 Dans le cas $p = q = \frac{1}{2}$, on a $g(a, s) = 1 - r_a = a/s$, qui, sous la condition $s \geq a + 1$, est maximum pour $s = a + 1$. Dans le cas $p \neq q$, on a : $g(a, s) = 1 - r_a = (1 - c^a)/(1 - c^s)$, avec $c := q/p$. Lorsque $0 < c < 1$, la fonction $s \mapsto 1/(1 - c^s)$ est décroissante, ainsi que $g(a, s)$, puisque $1 - c^a > 0$. Par conséquent, $g(a, s)$ prend sa valeur maximum pour $s = a + 1$. Lorsque $c > 1$, la fonction $s \mapsto 1/(1 - c^s)$ est croissante, mais comme $1 - c^a < 0$, la fonction $g(a, s)$ est décroissante et on a la même conclusion.

9.3 On a $G'(x) = h(x)\ln\rho/(1 - \rho^{sx})^2$, avec $h(x) = (a - s)\rho^{(s+a)x} - a\rho^{ax} - s\rho^{sx}$. Posons $p := (s - a)/s$, $q := a/s$. Alors $-h(x)/s = p\rho^{(s+a)x} + q\rho^{ax} - \rho^{p(s+a)x+qax}$. Le second membre est positif, puisque la fonction ρ^x est convexe. Donc $-h(x) > 0$ et $G'(x) < 0$. La fonction $G(x)$ est une fonction décroissante de x.

9.4 **(a)** Il suffit de faire dans les formules (9.3.3) et (93.6) les substitutions : $a \leftarrow x - c$ et $b \leftarrow d - x$. Pour obtenir $\mathrm{P}\{T_d^x < T_c^x\}$, il suffit, dans $\mathrm{P}\{T_c^x < T_d^x\}$ de faire les substitutions : $p \leftarrow q$, $q \leftarrow p$ et $(d - x) \leftarrow (x - c)$.

(b) En conditionnant par rapport à Y_1, il vient :
$\mathrm{P}\{\mu_x < +\infty\} = \mathrm{P}\{Y_1 = 1, \mu_x < +\infty\} + \mathrm{P}\{Y_1 = -1, \mu_x < +\infty\} =$
$\mathrm{P}\{Y_1 = 1\}\mathrm{P}\{\mu_x < +\infty \,|\, Y_1 = 1\} + \mathrm{P}\{Y_1 = -1\}\mathrm{P}\{\mu_x < +\infty \,|\, Y_1 = -1\} =$

$p\,\mathrm{P}\{T_x^{x+1} < +\infty\} + q\,\mathrm{P}\{T_x^{x-1} < +\infty\}$. En utilisant, selon les cas, les formules obtenues dans (a), on trouve : si $p = q = \frac{1}{2}$, alors $p \cdot 1 + q \cdot 1 = 1$; si $p/q < 1$, alors $p \cdot 1 + q(p/q) = 2p$; si $p/q > 1$, alors $p(q/p) + q \cdot 1 = 2q$. Dans tous les cas, $2\min\{p, q\}$.

(c) Supposons $p/q < 1$. Alors pour tout $t \geq 0$, on a $\mathrm{P}\{M \geq t\} = \mathrm{P}\{T_t^0 < +\infty\} = (p/q)^t$, qui est la fonction de survie d'une loi géométrique de paramètre q/p, sur tout $\mathbb{N}$, puisque $\mathrm{P}\{M \geq 0\} = 1$. Si $p/q > 1$, pour tout $t \leq 0$, on a $\mathrm{P}\{m \leq t\} = \mathrm{P}\{T_t^0 < +\infty\} = (q/p)^{-t}$. Comme $\mathrm{P}\{m \leq 0\} = 1$, la variable aléatoire m suit une loi géométrique de paramètre p/q sur $-\mathbb{N} = \{0, -1, -2, \dots\}$.

9.5 **(a)** Si le joueur a perdu les n premières parties, il a misé $-1 - 2 - 2^2 - \cdots - 2^{n-1} = -2^n + 1$, somme qu'il a perdue, donc $X_n = -2^n + 1$. Cet évènement a la probabilité $1/2^n$. L'évènement contraire est $\{X_n = 1\}$ et il a donc la probabilité : $1 - 1/2^n$. Ainsi, X_n ne prend que deux valeurs et $\mathbb{E}[X_n] = (-2^n + 1)(1/2^n) + 1 \cdot (1 - 1/2^n) = 0$.

(b) Pour $n \geq 1$, on a $\mathbb{E}[|X_n|] = (2^n - 1)(1/2^n) + (1 - 1/2^n) \leq 2$. Ensuite, $\mathbb{E}[X_{n+1} \,|\, X_0, X_1, \dots, X_n] = \mathbb{E}[X_{n+1} \,|\, X_n] = (X_n + 2^n)\frac{1}{2} + (X_n - 2^n)\frac{1}{2} = X_n$.

(c) Comme $\{T = n\} = \{X_1 < 1, \dots, X_{n-1} < 1, X_n = 1\}$ $(n \geq 1)$, la variable aléatoire T est bien un temps d'arrêt. Pour $n \geq 1$, on a donc $\mathrm{P}\{T = n\} = (1/2)^{n-1}(1/2) = (1/2)^n$ et T a ainsi une loi géométrique de paramètre $\frac{1}{2}$ sur $\{1, 2, \dots\}$. De plus, $\mathrm{P}\{T < +\infty\} = \sum_{n \geq 1} (1/2)^n = 1$. Comme $\mathbb{E}[X_T] = 1$ et $\mathbb{E}[X_n] = 0$ pour tout $n \geq 0$, le théorème d'arrêt n'est évidemment pas vérifié. Le temps d'arrêt T est presque sûrement fini. Donc, la seconde hypothèse du théorème d'arrêt, l'inégalité $\mathbb{E}[\sup_{n \geq 0} |X_{T \wedge n}|] < +\infty$ n'est pas satisfaite. Le processus arrêté $X^{|T} = (X_{T \wedge n})$ $(n \geq 0)$ n'est donc pas borné. Cela veut dire qu'avant de gagner 1 euro, le joueur peut perdre une somme illimitée. Il faut donc qu'il dispose de ressources illimitées à la fois en argent et en temps. Cette stratégie n'est donc pas recommandée !

(d) $\mathbb{E}[X_{n+1} \,|\, X_0, \dots, X_n] = \mathbb{E}[X_{n+1} \,|\, X_n] = (X_n + 2^n)p + (X_n - 2^n)q = X_n + (p - q)2^n$. Cette quantité est au plus égale ou au moins égale à X_n, suivant que $p < \frac{1}{2}$ ou $p > \frac{1}{2}$.

CHAPITRE 10

10.1 Noter que la suite $(|X_n|^p)$ $(n \geq 0)$ est une sous-martingale positive (*cf.* Lemme 7.4.2) et appliquer le Théorème 10.2.1.

10.2 Pour tout $c > 0$, la suite $((X_n + c)^2)$ $(n \geq 0)$ est une sous-martingale positive. On applique le Théorème 2.1 en notant que l'évènement $\{\max_{0 \leq k \leq n} X_k > \lambda\}$ entraîne $\{\max_{0 \leq k \leq n} (X_k + c)^2 > (\lambda + c)^2\}$, pour obtenir $\mathrm{P}\{\max_{0 \leq k \leq n} X_k > \lambda\} \leq \mathbb{E}[(X_n + c)^2]/(\lambda + c)^2$, une majoration qu'on pose égale à $g(c)$. On cherche une

valeur de $c > 0$ telle que $g(c)$ soit minimum et on trouve $c = \mathbb{E}[X_n^2]/\lambda$. En reportant dans $g(c)$, on trouve la majoration proposée.

10.3 **(a)** Comme S_{n+1} est une fonction de H_{n+1} (c'est sa première composante!), on a : $S_{n+1} = \mathbb{E}[S_{n+1} \mid H_{n+1}] = \sum_{k=1}^{n+1} \mathbb{E}[X_k \mid H_{n+1}] = (n+1) \times \mathbb{E}[X_1 \mid H_{n+1}]$. D'où $Y_{n+1} = \mathbb{E}[X_1 \mid H_{n+1}] = (1/n) \sum_{k=1}^{n} \mathbb{E}[X_1 \mid H_{n+1}] = (1/n) \sum_{k=1}^{n} \mathbb{E}[X_k \mid H_{n+1}] = \mathbb{E}[Y_n \mid H_{n+1}]$.

(b) La suite $(|Y_n|)$ est une sous-martingale renversée, *positive*, par rapport à la suite (H_n). Il en résulte que la suite $\mathbb{E}[|Y_n|])$ est croissante, lorsque n *décroît* vers 1, donc la suite $\mathbb{E}[|Y_n|])$ est décroissante, lorsque n tend vers l'infini. On peut appliquer le Théorème 10.3.2 : pour $1 \le n \le m$ et tout $\lambda > 0$, on a l'inégalité : $\lambda \mathrm{P}\{\min_{n\le k\le m} |Y_k| > \lambda\} \le \mathbb{E}[|Y_n|]$. La majoration étant uniforme en m, on en déduit l'inégalité cherchée.

10.4 Posons $U := \max_{0\le k\le n} X_k$. Alors $\mathbb{E}[U] = A + B$, avec $A := \int_0^x \mathrm{P}\{U > t\}\,dt$ et $B := \int_x^{+\infty} \mathrm{P}\{U > t\}\,dt$. D'abord, $A \le x$; on applique ensuite à $B = \int_x^{+\infty} \mathrm{P}\{U^p > t^p\}\,dt = \int_x^{+\infty} \mathrm{P}\{\max_{0\le k\le n} X_k^p > t\}\,dt$ l'inégalité (2.1) : $B \le \int_x^{+\infty} \mathbb{E}[X_n^p]/t^p\,dt = \mathbb{E}[X_n^p]\int_x^{+\infty}(1/t^p)\,dt = \mathbb{E}[X_n^p]\,x^{-(p-1)}/(p-1)$. D'où, $\mathbb{E}[U] \le g(x)$ avec $g(x) := x + \mathbb{E}[X_n^p]\,x^{-(p-1)}/(p-1)$. On cherche une valeur de $x > 0$ qui rend $g(x)$ minimum et on trouve $x = (\mathbb{E}[X_n^p])^{1/p} = \|X_n\|_p$. La valeur correspondante de $g(x)$ est donc $g(\|X_n\|_p) = \|X_n\|_p + \mathbb{E}[X_n^p]\,(\mathbb{E}[X_n^p])^{-(p-1)/p}/(p-1) = (p/(p-1))\,\|X_n\|_p$.

10.5 La suite (S_n) $(n \ge 0)$ est une martingale adaptée à (Y_n) $(n \ge 0)$. Donc la suite (S_n^2) $(n \ge 0)$ est une sous-martingale adaptée à (Y_n) $(n \ge 0)$. On peut donc lui appliquer l'inégalité maximale (3.1). En posant $M_n := \{\max_{0\le k\le n} S_k^2 > \lambda\}$, on sait que, pour tout $\lambda > 0$, on a $\lambda \mathrm{P}(M_n) \le \mathbb{E}[S_n^2\, I_{M_n}]$. En posant $\lambda := \varepsilon^2$, $\varepsilon > 0$, on a $M_n := \{\max_{0\le k\le n} |S_k| > \varepsilon\}$ et l'on obtient bien l'inégalité cherchée.

10.6 Si $0 \le a < 1$, il s'agit de démontrer l'inégalité : $a \operatorname{Log} b \le \frac{b}{e}$. Celle-ci est banale pour $0 < b < 1$. Pour $b \ge 1$, il suffit de démontrer : $\frac{\operatorname{Log} b}{b} \le \frac{1}{ae}$. Or, la fonction $\frac{\operatorname{Log} x}{x}$ prend son maximum pour $x = e$; d'où $\frac{\operatorname{Log} b}{b} \le \frac{1}{e} \le \frac{1}{ae}$. Si $a \ge 1$, il s'agit de démontrer l'inégalité $a \operatorname{Log} b \le a \operatorname{Log} a + \frac{b}{e}$, c'est-à-dire $\frac{a}{b} \operatorname{Log} \frac{b}{a} \le \frac{1}{e}$, qui est satisfaite, puisque la fonction $\frac{\operatorname{Log} x}{x}$ prend son maximum en $x = e$.

CHAPITRE 11

11.1 En effet, pour chaque $n \geq 0$, on a $\mathbb{E}[|X_n|] = \mathbb{E}[|\mathbb{E}[Z \mid Y_0, \dots, Y_n]|] \leq \mathbb{E}[\mathbb{E}[|Z| \mid Y_0, \dots, Y_n]] = \mathbb{E}[|Z|] < +\infty$ par hypothèse. Le Théorème 11.1.1 s'applique : $X_n \xrightarrow{p.s.} X_\infty$.

11.2 Pour chaque $n \geq 0$, posons : $Z_n := \mathbb{E}[Z \mid Y_0, \dots, Y_n]$. Comme $X_n \leq Z$ et que (X_n) est une martingale adaptée par rapport à (Y_n), on a : $X_n = \mathbb{E}[X_n \mid Y_0, \dots, Y_n] \leq \mathbb{E}[Z \mid Y_0, \dots, Y_n] = Z_n$. Or, (Z_n) $(n \geq 0)$ est une martingale de Doob, qui converge presque sûrement, d'après l'Exercice 1. D'autre part, $(X_n - Z_n)$ $(n \geq 0)$, différence de deux martingales, est encore une martingale, telle que pour tout $n \geq 0$ on a $X_n - Z_n \leq 0$. D'après le Corollaire 11.1.5, elle converge presque sûrement. Donc la suite de terme général $X_n = Z_n + (X_n - Z_n)$ converge aussi presque sûrement.

11.3 Pour la martingale arrêtée $X^{|n} = (X_{m \wedge n})$ $(m \geq 0)$, on a $\{S_k^{|n} \leq n\} = \{S_k < +\infty\}$ et $\{T_k^{|n} \leq n\} = \{T_k < +\infty\}$. L'inégalité de Dubins donne :

$$\forall k \geq 1 \quad (b-a)\,\mathrm{P}\{T_k^{|n} < n\} \leq \mathbb{E}[(a - X_n)\, I_{\{S_k^{|n} \leq n < T_k^{|n}\}}];$$

soit $\forall k \geq 1 \quad (b-a)\,\mathrm{P}\{T_k < +\infty\} \leq \mathbb{E}[(a - X_n)\, I_{\{S_k < +\infty, T_k = +\infty\}}]$; d'où l'inégalité à prouver, puisque $\mathrm{P}\{M \geq k\} = \mathrm{P}\{T_k < +\infty\}$.

11.4 **(a)** Pour tout $n \geq 1$, la variable aléatoire X_n est du premier ordre; de plus, on obtient : $\mathbb{E}[X_n] = \prod_{k=1}^{n} \mathbb{E}[Y_k] = 1$; enfin, $\mathbb{E}[X_{n+1} \mid Y_1, \dots, Y_n] = \mathbb{E}[X_n Y_{n+1} \mid Y_1, \dots, Y_n] = X_n \mathbb{E}[Y_{n+1} \mid Y_1, \dots, Y_n] = X_n \mathbb{E}[Y_{n+1}] = X_n$.

(b) Pour le cas $p := \mathrm{P}\{Y_1 = 0\} > 0$, on part de l'inclusion $\bigcup_{k \geq 1} \{Y_k = 0\} \subset \{X_n \to 0\}$. Il suffit alors de montrer que la probabilité du membre de gauche de cette inclusion est égal à 1. Or, $\mathrm{P}(\bigcup_{k \geq 1} \{Y_k = 0\}) = \mathrm{P}(\lim_n \bigcup_{k=1}^{n} \{Y_k = 0\}) = \lim_n \mathrm{P}(\bigcup_{k=1}^{n} \{Y_k = 0\}) = \lim_n (1 - \mathrm{P}(\bigcap_{k=1}^{n} \{Y_k \neq 0\})) = \lim_n (1 - \prod_{k=1}^{n} \mathrm{P}\{Y_k \neq 0\}) = \lim_n (1 - (1-p)^n) = 1$, puisque $0 \leq 1 - p < 1$.

Dans le cas $p = \mathrm{P}\{Y_1 = 0\} = 0$, les variables aléatoires $\operatorname{Log} Y_n$ sont presque sûrement définies. Il en est donc de même des variables $\operatorname{Log} X_n = \sum_{k=1}^{n} \operatorname{Log} Y_k$. Or, la fonction $x \mapsto -\operatorname{Log} x$ est *concave*. D'après l'inégalité de Jensen, on a : $0 = -\operatorname{Log} \mathbb{E}[Y_1] \leq -\mathbb{E}[\operatorname{Log} Y_1]$. D'autre part, la loi de Y_1 n'étant pas dégénérée, on a l'inégalité *stricte* $\mathbb{E}[\operatorname{Log} Y_1] < 0$. Appliquons la loi forte des grands nombres : $\frac{1}{n} \operatorname{Log} X_n = \frac{1}{n} \sum_{k=1}^{n} \operatorname{Log} Y_k \xrightarrow{p.s.} \mathbb{E}[\operatorname{Log} Y_1] < 0$, d'où $\operatorname{Log} X_n \xrightarrow{p.s.} -\infty$ et par conséquent $X_n \xrightarrow{p.s.} 0$.

Remarque 1 : Le seul cas où X_n ne converge pas vers 0 est celui où la loi de Y_1 est dégénérée, c'est-à-dire $Y_1 = 1$ p.s. Pour tout $n \geq 1$, on a alors $X_n = 1$ p.s. pour tout $n \geq 1$, d'où $X_n \xrightarrow{p.s.} 1$, lorsque n tend vers l'infini.

Remarque 2 : La suite (X_n) $(n \geq 1)$ est un exemple d'une suite de variables aléatoires *positives*, toutes d'espérance mathématique finie et égale à 1 et pourtant telles que $X_n \xrightarrow{p.s.} 0$, lorsque n tend vers l'infini.

Remarque 3 : Si les variables aléatoires ne sont plus supposées identiquement distribuées, le présent exercice donne lieu au théorème des martingales "produit" de Kakutani, dont le traitement n'est pas aussi élémentaire. Voir le livre de Williams.[1]

11.5 (a) A la date n, l'urne contient $(n+2)$ boules, dont N_n blanches. La quantité X_n représente donc le rapport, à la date n, du nombre de boules blanches dans l'urne au nombre total de boules. D'abord, $\mathbb{E}[\,|X_n|\,] = \mathbb{E}[X_n] \leq 1 < +\infty$ pour tout $n \geq 0$. Ensuite, conditionnellement à $(N_0, N_1, \dots, N_n)$ (en fait, conditionnellement à N_n), la probabilité d'amener une boule blanche à la $(n+1)^{\text{ième}}$ opération est $N_n/(n+2)$, d'où $\mathbb{E}[N_{n+1} \,|\, N_0, N_1, \dots, N_n] = N_n + (N_n/(n+2)) = (n+3)N_n/(n+2)$, soit $\mathbb{E}[X_{n+1} \,|\, N_0, N_1, \dots, N_n] = X_n$.

(b) La martingale (X_n) $(n \geq 0)$ est bornée, puisque $0 \leq X_n \leq 1$ pour tout $n \geq 0$. Elle converge donc presque sûrement vers une variable aléatoire X intégrable. Le rapport du nombre de boules blanches au nombre de boules noires après la $n^{\text{ième}}$ opération est $N_n/((n+2) - N_n) = X_n/(1 - X_n)$, une quantité qui tend presque sûrement vers $X/(1-X)$ si $X \neq 1$. [Le cas $X = 1$ peut, par exemple, se produire, lorsque l'on tire une boule blanche à chaque opération. Alors $X_n = (n+1)/(n+2)$ tend vers $X = 1$ et le rapport $X_n/(1 - X_n) = n + 1$ tend vers $+\infty$.]

CHAPITRE 13

13.1 (a) On a $\operatorname{Cov}(X(s), X(t)) = \mathbb{E}[X(s)\,X(t)] = \mathbb{E}[X(s)(X(s) + (X(t) - X(s)))] = \mathbb{E}[X^2(s)] + \mathbb{E}[X(s)(X(t) - X(s))]$. Or, $X(s)$ et $X(t) - X(s)$ sont indépendants. On en tire : $\operatorname{Cov}(X(s), X(t)) = \mathbb{E}[X^2(s)] + \mathbb{E}[X(s)] \times \mathbb{E}[X(t) - X(s)] = \mathbb{E}[X^2(s)] = \operatorname{Var} X(s) = s$.

(b) La densité de $X(s)$ conditionnellement à $\{X(t) = b\}$ est le rapport A/B de la densité conjointe A de $(X(s), X(t))$ au point (x, b) à la densité B de $X(t)$ au point b. Or, la densité de $(X(s), X(t))$ au point (x, b) est la même que la densité de $(X(s), X(t) - X(s))$ au point $(x, b - x)$. Ces deux variables étant indépendantes, on a donc :

$$A = \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{s}} \, e^{-x^2/(2s)} \cdot \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{t-s}} \, e^{-(b-x)^2/(2(t-s))};$$

---

[1] Williams (David). — *Probability with martingales*, Cambridge Univ. Press, 1991, p. 144.

$$B = \frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{t}}\, e^{-b^2/(2t)}; \quad \text{soit} \quad \frac{A}{B} = \frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{s\left(1-\frac{s}{t}\right)}} \exp\left(-\frac{1}{2}\,\frac{\left(x-\frac{s}{t}b\right)^2}{s\left(1-\frac{s}{t}\right)}\right).$$

(c) D'abord, $\mathrm{Cov}((X(s), X(t)) \mid X(1) = 0) = \mathbb{E}[X(s)X(t) \mid X(1) = 0]$. Puis, $\mathbb{E}[X(s)X(t) \mid X(1)] = \mathbb{E}_{X(1)}[\mathbb{E}_{X(1)}[X(s)X(t) \mid X(t)]$, soit $\mathbb{E}[X(s)X(t) \mid X(1) = 0] = \mathbb{E}_{\{X(1)=0\}}[\mathbb{E}_{\{X(1)=0\}}[X(s)X(t) \mid X(t)] =$ $\mathbb{E}_{\{X(1)=0\}}[X(t)\,\mathbb{E}_{\{X(1)=0\}}[X(s) \mid X(t)] = \mathbb{E}_{\{X(1)=0\}}[X(t)\,(s/t)X(t)] =$ $(s/t)\ \mathrm{Var}(X(t) \mid X(1) = 0) = (s/t)\,t(1-t) = s(1-t)$.

13.2 $\mathrm{P}\{\inf_{0\le s\le t} X(s) \le 0 \mid X(0) = a\} = \mathrm{P}\{\sup_{0\le s\le t} X(s) \ge 0 \mid X(0) = -a\}$ par symétrie. A cause de la propriété des accroissements stationnaires, cette dernière quantité est égale à $\mathrm{P}\{\sup_{0\le s\le t} X(s) \ge 0 \mid X(0) = 0\}$, qui vaut encore $\mathrm{P}\{T_a \le t\} = (a/\sqrt{2\pi}) \int_0^t u^{-3/2} e^{-a^2/(2u)}\, du$.

13.3 Pour $a > 0$, la formule résulte de l'exercice précédent, en vertu de la propriété des accroissements stationnaires.

13.4 On a $\mathrm{P}(B) = \int_{\mathbb{R}} \mathrm{P}\{B \mid |X(t_0)| = a\}\, f_{|X(t_0)|}(a)\, da$, avec $f_{|X(t_0)|}(a) = \sqrt{2/(\pi t_0)}\, e^{-a^2/(2t_0)}$ pour $a > 0$ et 0 autrement et où $\mathrm{P}\{B \mid |X(t_0)| = a\}$ n'est autre que la probabilité $P(a)$ évaluée dans l'Exercice 3. On en tire :

$$\begin{aligned}
\alpha &= \sqrt{\frac{2}{\pi t_0}} \int_0^{+\infty} P(a)\, e^{-a^2/(2t_0)}\, da \\
&= \sqrt{\frac{2}{\pi t_0}} \int_0^{+\infty} e^{-a^2/(2t_0)} \Big(\frac{a}{\sqrt{2\pi}} \int_0^{t_1-t_0} u^{-3/2} e^{-a^2/(2u)}\, du\Big) da \\
&= \frac{1}{\pi\sqrt{t_0}} \int_0^{t_1-t_0} u^{-3/2}\, du \int_0^{+\infty} a\, e^{(-a^2/2)((1/u)+(1/t_0))}\, da.
\end{aligned}$$

Avec le changement de variables $a^2/2 := x$, la dernière intégrale vaut $1/((1/u) + (1/t_0)) = t_0 u/(t_0 + u)$. D'où, avec le changement de variables $u := t_0 v^2$ $(v > 0)$,

$$\alpha = \frac{\sqrt{t_0}}{\pi} \int_0^{t_1-t_0} \frac{du}{(t_0+u)\sqrt{u}} = \frac{2}{\pi} \int_0^{\sqrt{(t_1-t_0)/t_0}} \frac{dv}{1+v^2} = \frac{2}{\pi} \mathrm{Arctg} \sqrt{\frac{t_1-t_0}{t_0}}.$$

On en tire :

$$\mathrm{tg}(\pi\alpha/2) = \sqrt{\frac{t_1-t_0}{t_0}} = \sqrt{\frac{1-(t_0/t_1)}{t_0/t_1}}.$$

D'où $\sqrt{\dfrac{t_0}{t_1}} = \cos(\pi\alpha/2)$ et $\alpha = \dfrac{2}{\pi}\,\mathrm{Arc}\cos\sqrt{\dfrac{t_0}{t_1}}$.

13.5 Pour $i = 1, 2$ on définit $B_i(t)$ par $X_i(t) =: x_i + B_i(t)$. Alors $\{B(t) : t \geq 0\}$, où $B(t) := B_1(t) - B_2(t)$, est un mouvement brownien tel que $B(t)$ est d'espérance $(\mu_1 - \mu_2)t$ et de variance $(\sigma_1^2 + \sigma_2^2)t$. De plus, $X_1(t) - X_2(t) = (x_1 - x_2) + B(t)$ et $B(0) = 0$. Soit T_c^0 le temps d'atteinte de c à partir de 0 pour le processus $\{B(t) : t \geq 0\}$. Les deux processus ne se rencontrent jamais si et seulement si le processus $\{B(t) : t \geq 0\}$, partant de 0, n'atteint jamais la position $-(x_1 - x_2) = x_2 - x_1$, c'est-à-dire si l'évènement $\{T^0_{x_2 - x_1} = +\infty\}$ est réalisé. Comme $\mathrm{P}\{T^0_{x_2-x_1} = +\infty\} = 1 - \mathrm{P}\{T^0_{x_2-x_1} < +\infty\}$ et $x_2 - x_1 > 0$, on obtient, d'après (7.5) et (7.6),

$$\mathrm{P}\{T^0_{x_2-x_1} = +\infty\} = \begin{cases} 0, & \text{si } \mu_1 - \mu_2 = 0\,; \\ 0, & \text{si } \mu_1 - \mu_2 > 0\,; \\ 1 - e^{-2(\mu_2 - \mu_1)(x_2 - x_1)/(\sigma_1^2 + \sigma_2^2)}, & \text{si } \mu_1 - \mu_2 < 0. \end{cases}$$

13.6 Supposons d'abord $\mu > 0$. Si $x < 0$, alors $\mathrm{P}\{Y > x\} = 1$, car Y est positive. Si $x > 0$, on a $\{Y > x\} = \{\inf\limits_{t>0} X(t) < -x\} = \{T^0_{-x} < +\infty\}$, d'où, d'après (13.7.5), $\mathrm{P}\{Y > x\} = \mathrm{P}\{T^0_{-x} < +\infty\} = e^{-2\mu x/\sigma^2}$. La fonction de survie de Y est celle d'une variable exponentielle. Supposons ensuite $\mu < 0$. Si $x < 0$, on a $\mathrm{P}\{Z > x\} = 1$, car Z est positive. Si $x > 0$, on a $\{Z > x\} = \{\sup\limits_{t>0} X(t) > x\} = \{T^0_x < +\infty\}$, d'où, d'après (13.7.6), $\mathrm{P}\{Z > x\} = \mathrm{P}\{T^0_x < +\infty\} = e^{-2|\mu| x/\sigma^2}$. La fonction de survie de Z est celle d'une variable exponentielle. On voit que si $\mu x < 0$, on a $\mathrm{P}\{T^0_x < +\infty\} = e^{2\mu x/\sigma^2} < 1$ et que si $\mu x > 0$ ou si $\mu = 0$ et x quelconque, on a $\mathrm{P}\{T^0_x < +\infty\} = 1$.

13.7 **(a)** On écrit $\mathbb{E}[Y' \, {}^tZ'] = 0$, soit $\mathbb{E}[(Y - A\,Z)\, {}^tZ'] = \mathbb{E}[Y\, {}^tZ'] - A\, \mathbb{E}[Z\, {}^tZ'] = \Gamma_{12} - A\,\Gamma_{22} = 0$.

(b) Comme Y' et Z' suivent des lois normales et sont non-corrélés, ils sont nécessairement indépendants. La matrice des variances-covariances de Y' vaut : $\Gamma' = \mathbb{E}[Y' \, {}^tY'] = \Gamma_{11} - \Gamma_{12}\,\Gamma_{22}^{-1}\,\Gamma_{21}$.

(c) Dn a : $\mathrm{P}\{\bigcap\limits_{i=1}^{p} \{Y_i \in B_i\} \,|\, Z = z\} = \mathrm{P}\{\bigcap\limits_{i=1}^{p} \{Y'_i + A\,Z_i \in B_i\}\} = \mathrm{P}\{\bigcap\limits_{i=1}^{p} \{Y'_i + A\,z_i \in B_i\}\}$.

13.8 **(a)** Une manière de procéder est de dériver l'expression intégrale de $I(t)$ sous le signe somme (la justification est aisée) et d'obtenir l'équation différentielle : $I'(t) = 0$; on prend alors la limite de $I(t)$ lorsque t tend vers $0+$. On peut aussi remarquer que si f est une fonction continue sur toute la droite (sauf éventuellement en 0), paire et intégrable, on a : $\int_{\mathbb{R}} f(u - t/u)\, du = \int_0^{+\infty} f(u - t/u)\, du + \int_{-\infty}^{0} f(u - t/u)\, du = 2\int_{\mathbb{R}} f(v)(1/2 + v(v^2 + 4t)^{-1/2})\, dv$, par le changement de variables $v = u - t/u$. Comme f est paire, on en déduit : $\int_{\mathbb{R}} f(u - t/u)\, du = 2\int_0^{+\infty} f(u - t/u)\, du = 2\int_0^{+\infty} f(v)\, dv$. La seconde intégrale se déduit immédiatement de la première.

(b) Prenons $t := x\sqrt{s}$ dans l'expression (β) de $J(t)$. On obtient :

$$\frac{2}{\sqrt{2\pi}}\int_0^{+\infty}\exp\Bigl(-\frac{1}{2}\Bigl(u^2+\frac{(x\sqrt{s})^2}{u^2}\Bigr)\Bigr)\,du = e^{-x\sqrt{s}}.$$

En faisant le changement de variables $t = x^2/(2u^2)$, il vient

$$\int_0^{+\infty} e^{-st}\frac{x}{2\sqrt{\pi}}t^{-3/2}e^{-x^2/(4t)}\,dt = e^{-x\sqrt{s}}.$$

C'est l'expression de (b).

(c) L'identité obtenue en (b) se récrit

$$\mathcal{L}^{-1}\,e^{-x\sqrt{s}} = (x/2\sqrt{\pi})t^{-3/2}e^{-x^2/(4t)},$$

une expression que l'on pose égale à $f(t)$. On en tire :
$\mathcal{L}^{-1}\,e^{-(\sqrt{\mu^2+2s\sigma^2}-\mu)(x/\sigma^2)} = e^{x\mu/\sigma^2}\mathcal{L}^{-1}\,e^{-\sqrt{\mu^2+2s\sigma^2}(x/\sigma^2)} =$
$e^{x\mu/\sigma^2}\mathcal{L}^{-1}\,e^{-\sqrt{2\sigma^2(s+\mu^2/(2\sigma^2))}(x/\sigma^2)} = e^{x\mu/\sigma^2}e^{-\mu^2t/(2\sigma^2)}\mathcal{L}^{-1}\,e^{-\sqrt{2\sigma^2 s}(x/\sigma^2)} =$
$e^{x\mu/\sigma^2}e^{-\mu^2t/(2\sigma^2)}\mathcal{L}^{-1}\,e^{-x\sqrt{s/(\sigma^2/2)}} = e^{x\mu/\sigma^2}e^{-\mu^2t/(2\sigma^2)}(\sigma^2/2)\,f(t\sigma^2/2) =$
$e^{x\mu/\sigma^2}e^{-\mu^2t/(2\sigma^2)}(\sigma^2/2)\,(x/2\sqrt{\pi})t^{-3/2}(\sigma^2/2)^{-3/2}e^{-x^2/(4t\sigma^2/2)}$, soit l'expression de $f_{T_x^0}(t)$ de (13.8.4) :

$$f_{T_x^0}(t) = \frac{1}{\sqrt{2\pi}}\frac{x}{\sigma}t^{-3/2}\exp\Bigl(-\frac{1}{2\sigma^2}\Bigl(\frac{x}{\sqrt{t}}-\mu\sqrt{t}\Bigr)^2\Bigr).$$

13.9 Pour tout entier $k \ge 1$, l'évènement $\{|X(1)| \le 1/k\}$ est réalisé si et seulement si l'évènement $\bigcap_{i=1}^n \{|Z_i| \le t_i/k\}$ est réalisé. Pour k assez grand, on a la double inclusion : $\bigcap_{i=1}^n \{a_i + 1/k < Y_i \le b_i - 1/k, |Z_i| \le t_i/k\} \subset \bigcap_{i=1}^n \{a_i < Y_i + Z_i \le b_i, |Z_i| \le t_i/k\} \subset \bigcap_{i=1}^n \{a_i - 1/k < Y_i \le b_i + 1/k, |Z_i| \le t_i/k\}$. En passant aux probabilités, on obtient une double inégalité. Divisons alors les trois membres de celle-ci par $\mathrm{P}\{|X(1)| \le 1/k\}$. Comme chacun des évènements $\{|Z_i| \le t_i/k\}$ $(i = 1,\dots,n)$ est égal à $\{|X(1)| \le 1/k\}$ et que Y et Z sont indépendants, on obtient un encadrement pour la probabilité conditionnelle $\mathrm{P}\{\bigcap_{i=1}^n \{a_i < Y_i + Z_i \le b_i\}\,|\,|X(1)| \le 1/k\}$ de la forme : $\mathrm{P}\{\bigcap_{i=1}^n \{a_i + 1/k < Y_i \le b_i - 1/k\}\} \le \mathrm{P}\{\bigcap_{i=1}^n \{a_i < Y_i + Z_i \le b_i\} \bigm| |X(1)| \le 1/k\} \le \mathrm{P}\{\bigcap_{i=1}^n \{a_i - 1/k < Y_i \le b_i + 1/k\}\}$. En faisant tendre k vers l'infini, on en déduit : $\mathrm{P}\{\bigcap_{i=1}^n \{a_i < Y_i \le b_i\} = \mathrm{P}\{\bigcap_{i=1}^n \{a_i < Y_i + Z_i \le b_i\} \bigm| |X(1)| = 0\} = \mathrm{P}\{\bigcap_{i=1}^n \{a_i < X(t_i) \le b_i\} \bigm| |X(1)| = 0\}$.

Index

D

E

K

L

M

N

O

P

R

S

048850 - (III) - (0,4) - OSB 80° - AUT - JME

Achevé d'imprimer sur les presses de
JOUVE
1, rue du Docteur Sauvé, 53100 Mayenne
Dépôt légal 1ère édition : avril 2002 - N° 529882K

Dépôt légal : octobre 2004 - suite du tirage : juillet 2010

Imprimé en France

SCIENCES SUP



Dominique Foata
Aimé Fuchs

PROCESSUS STOCHASTIQUES

Processus de Poisson, chaînes de Markov et martingales



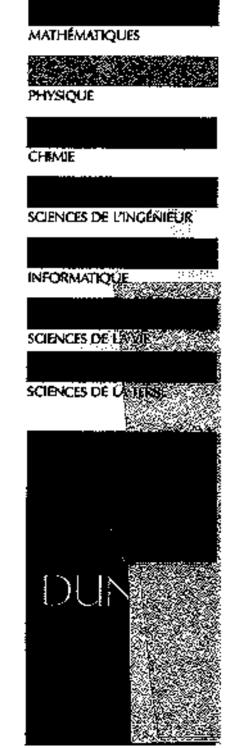
Ce livre s'adresse aux étudiants en 2ᵉ cycle/Master de mathématiques appliquées et d'informatique, ainsi qu'aux élèves des grandes écoles d'ingénieurs, qui s'orientent vers la recherche opérationnelle.

Il présuppose la connaissance d'un cours de probabilités de base, comme celui qui est exposé dans le livre *Calcul des probabilités*, écrit par les même auteurs.

On y trouve un exposé sur les processus de Poisson, les chaînes de Markov et les martingales à temps discret, ainsi qu'une brève introduction au mouvement brownien. Le livre comporte de nombreux exercices, dont la solution est généralement détaillée et un chapitre d'exemples d'applications, dans lesquels les différents processus sont utilisés.

DOMINIQUE FOATA et AIMÉ FUCHS

sont professeurs à l'université Louis-Pasteur de Strasbourg. Ils ont publié séparément plusieurs monographies et articles, dans le domaine de la combinatoire classique et algébrique et des fonctions spéciales pour Dominique Foata, et dans celui des probabilités et de la statistique pour Aimé Fuchs.







ISBN 2 10 048850 3

www.dunod.com